

OM of: US-09-494-297-2 to: N\_Geneseq\_0401:\* out\_format : pfs  
Date: Jun 7, 2001 12:15 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Comugen Ltd.

## Command line parameters:

```
-MODEL-frame-1p2n.model -DEV-xlp  
-o-/cgnl_1/USPTO_spool/US09494297/runat-06062001.115735.12347/app-query.fasta.1.825  
-db-N_Geneseq_0401 -QMT-fastap -SUFFIX-rng -GAPOP-12.000  
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPEXT-0.000 -LOOPEXT-0.000  
-GAPEXT-4.500 -GAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500  
-XGAPOP-6.000 -XGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500  
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-biosum62  
-TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR_SCORE-pct  
-THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL -OUTFMT=ps  
-NORM-ext -MINLEN=0 -MAXLEN=200000000  
-USER-US09494297_ecgnl_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPRX  
-WAIT -THREADS=1
```

## Search information block:

Query: US-09-494-297-2  
Query length: 757  
Database: N\_Geneseq\_0401:\*  
Database sequences: 678276  
Database length: 291890651  
Search time (sec): 149.980000

## score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X12982		196.00	241.91	1.6e-05	15614
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A30874		180.00	238.53	2.4e-05	2127
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:V82019		178.00	223.16	0.0002	7650
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:V82021		176.00	230.41	6.8e-05	2718
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20201		161.00	204.61	0.0019	4249
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20200		161.00	204.36	0.0019	4359
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X13179		161.00	194.10	0.0071	12445
/cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:T68844		154.00	192.00	0.0093	5547
/cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V29930		152.00	189.14	0.0135	5547
/cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:T68843		152.00	186.13	0.0198	7551
/cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V29928		152.00	186.13	0.0198	7551
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X77451		150.00	190.70	0.0110	3531
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A97541		149.00	191.89	0.0095	2700
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X77593		149.00	188.18	0.0152	3945
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20228		149.00	183.81	0.0267	6168
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20230		149.00	183.81	0.0267	6168
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20232		149.00	183.81	0.0267	6168
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X13101		148.00	169.25	0.1727	27360
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A68609		148.00	163.28	0.3719	43576
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X99560		147.00	172.26	0.1175	15016
/cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT:Q24123		146.00	182.37	0.0321	4612
/cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V39862		145.50	186.99	0.0178	2672
/cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V04279		144.50	182.65	0.0310	3600
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:Z93534		142.00	180.93	0.0386	2976
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:Z51202		142.00	180.93	0.0386	2976
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20209		139.00	174.15	0.0878	3695
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20208		139.00	174.15	0.0921	3840
/cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT:T29247		139.00	168.15	0.1991	7101
/cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V30561		139.00	168.15	0.1991	7101
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X13104		138.00	158.51	0.6853	19031
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:V82020		138.00	166.00	0.2623	7642
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X13026		137.50	155.85	0.9644	20072
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X206831		137.50	149.62	2.14	37948
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X13301		137.00	162.43	0.4143	9509
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:CA42710		136.00	176.54	0.0679	1940
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:CS0213		136.00	166.79	0.2369	5259
/cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:AV5565		135.50	172.74	0.1105	2661
/cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:T75487		135.00	168.84	0.1823	20986
/cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X22748		135.00	168.84	0.1823	3687

/cgnl\_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A71797 + 135.00 168.84 0.1823 3687  
/cgnl\_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X78245 + 135.00 168.62 0.1873 3768  
/cgnl\_8/gcgdata/geneseq/geneseqn/NA1996.DAT:T08695 + 135.00 168.46 0.1913 3832  
/cgnl\_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X78245 + 135.00 165.88 0.2663 4968  
/cgnl\_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X99555 - 134.00 169.24 0.1731 3057

seq\_name: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X12982

## seq\_documentation\_block:

ID X12982 standard; DNA; 15614 BP.

AC X12982;

DT 19-MAR-1999 (first entry)

XX Enterococcus faecalis genome contig SEQ ID NO:45.

XX Enterococcus faecalis; contig; detection; Enterococcal Infection;

KM vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

OS WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98MO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046555.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX Claim 1; Page 419-427; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

XX X12982 to X13919 represent these nucleotide sequences which are primary

XX nucleotide sequences, also known as contigs. The computer-based system

XX can identify fragments of the Enterococcus faecalis genome with

XX commercial importance. The products can be used to detect the presence

XX of Enterococcus faecalis in samples. They can also be used for

XX diagnosing Enterococcal infection in an animal and monitoring

XX progression of disease, and for identifying agents which can be used to

XX modulate the growth or pathogenicity of Enterococcus faecalis, or

XX another related organism, in vivo or in vitro. In particular the

XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences

XX can be used in vaccines to prevent or attenuate an Enterococcal

XX infection.

XX Sequence 15614 BP; 5535 A; 2567 C; 3169 G; 4326 T; 17 other;

XX

alignment\_scores: Quality: 196.00 Length: 706

Ratio: 0.596 Gaps: 42

Percent Similarity: 46.601 Percent Identity: 22.946

alignment\_block: US-09-494-297-2 x X12982 ..

Align seg 1/1 to: X12982 from: 1 to: 15614

```

96 ValAsnLeuGluGlySer.....ArgSerThrGlnValTyrCysPheAs 110
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9070 GTGTCGCTAAGCGCGCTACATAATATATAGATTTAT..... 9111
110 nLeuLysLysAlaPheProLeuGlySerAspSerSerValLysTyrTrp 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9112 .....TTGGGTAAAGCGCGCAAGAAATTCAAATTCATTT 9142
127 YrLysLysHisAspGlyLysSerThrLysPheGluAsp..... 139
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9143 ATCAAGTACGCT.....ATTCAAAACAGAGTCAGCAAAACTTCAAACCTGAT 9186
140 .....TyrAlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLys 154
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9187 TTTTGGATCAATGAATGCTCGGACCAACGTTTACGCCATTAGCCACGCC 9236
154 sLeuArgAlaValMetTyrAsnGlyHisProGlnAsnAlaAsnGlyLeu 171
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9237 CCCTGAAAAAGTGTGATTT.....GGGGTTC 9262
171 etGluGlyLeuGluProLeuAsnAlaIleArgValThrGlnGluAlaVal 187
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9263 CTTCGGGAAAAAGCACCT.....GGCGGAAGTTAAACGTGAAAAAAATC 9306
188 TrpTyrTyrSerAspAsnAlaProIleSerAsnProAspGluSerPheLys 204
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9307 TGGGAGAGTATGATCAAGACCGACCAAGTCGGCCAGTATATGATTTA 9356
204 sArgGluSerGluSerAsnLeuValSerThrSer..... 215
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9357 TGAATATAGTAAAGCAAGTAACTGACACACGCACTGCGCAACTGGGT 9406
216 .....GlnLeuSerLeuMetArgGlnAlaLeuLysGlnLeuIleAspPro 230
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9407 ATATTAATATTCAAAACCCAGAAAATGATACCAATAGTTGGGAGCGC 9456
231 AsnLeuAlaThrLysMetProLysGlnValProAspAspPheGln..... 245
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9457 AAAAATGTAACCCCACTTCCAAAACCGCGGATGAAGAAGCTATCAAGAAGT 9506
246 .LeuSerIlePheGluSerGluAspLysGlyAspLysTyrAsnLysGlyT 262
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9507 TCTTGGGCTTCCCAATACAAACAATCAAGCAAGCTTTCAT.....T 9550
262 YrGlnAsnLeuLeuSerGlyLysLeuValProThrLysProProThrPro 278
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9551 ATCAAAACACCCGTAATTAACA...GTTCTGTTACAGTCAAGAAAAA 9597
279 GlyAspProMetProProAsnGlnProGlnThrThrSerVal..... 293
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9598 ATGACGATTACTTGTGAAAAACACGACAGTTCAAGCCATTAGATTT 9647
294 ....LeuIleArgLysTyrAlaIleGlyAspTyrSerLysLeuLeuGlu 309
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9648 AAAAGTATATCAAAAATCTTCTCAGGTGAG.....AAAACCTTAGTGG 9691
309 LysAlaThrLeuGlnLeuThrGlyAspAsnValAsnSerPheGlnAlaArg 325
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9692 GAGCGCTTTGTAATGAGTGTGTAATAATGTT.....CAAAACACA 9732
326 ValPheSerSerAsnAsp.....IleGlyLysArgIleGluLeu 338
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9733 TTATGTGACATAAAGATGAGTATTCCTTGCACAAAGATGCGCGCT 9782
338 uSerAspGly...ThrTyrThrLeuThrGluLeuAsnSerProAlaGlyT 354
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9783 ACAAAAAGGGGAGACGCTATACATTAAGTAAAGACACCTCAGAGAC 9832
354 YrSerIleAlaGluProIleThrPheLysValGluAla.....Gly 367
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9833 ATGAGTTAGGCAAGAAAAAGCACTTGGCAAAATGAGGTGAGTGAACCAAGC 9882
368 LysValTyrThrIleIleAspGlyLysGlnIleGluAsnProAsnLysGln 384

```

```

9883 AAAGTA.....ACCATGATGACAGCAAGAGTGACACCAACAAATCAAGT 9926
384 uIleValGluProTyrSerValGluAlaTyrAsnAspPheGluGluPheS 401
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9927 TATTT.....CCATTGGAAATTTGAA.....AAATAAATTTCTTCTTGC 9964
401 erVal.....LeuThrGlnAsnTyrAlaLysPheTyrTyr 413
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9965 CATTCAGATTAAGAAATATACACCATGCAAAAT..... 9996
414 AlaLysAsnLysAsnGlySerSer.GlnValValTyrCysPheAsnAla 430
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9997 GGCAACAGAGTGAAGTATAGCAGAGCGCACCTTTCGTTGCAAGAA.... 10042
430 sPLeuLysSerProProAspSerGluAspGlyLysTyrMetThrPro 446
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10043 .....AAATGCTGCAAGCA 10057
447 AspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAspLe 463
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10058 AGTTACCAAACTGTGCAACTCAAAAAACAGATTAATCAGGATTGACCTA 10107
463 uPheLysTyrThrValLysProArgAspThrAspProAspThrPheLeuL 480
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10108 TTTT..... 10111
480 YrHisIleLysLysValIleGluLysGly...TyrArgGluLysGlyGln 495
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10112 .....AAATTAAGTGAACCTGTGATGATCA..... 10138
496 AlaIleGluTyrSerGlyLeuThrGluThrGlnLeuArgAlaIleThrGln 512
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10139 ATGTTGGACAAATCAGCA.....CCATTAGCTACGACACTCT 10176
512 nLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuAspLysLysL 529
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10177 TGTGCAAAATTAATGAATTTACT.....GTTGATAAA..... 10207
529 euLysAspTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaVal 545
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10208 .....TATGGAAA..... 10216
546 AlaLysIleLeuValGluTyrAlaGln.....AspSerAsnProPr 559
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10217 .....ATTCACTATGACGCAAAAAATTTCAAGAAAAATGCGCC 10254
559 OGlnLeuThrAspLeuAspPhePheIleProAsnAsnAsnLysTyrGlnS 576
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10255 AGAATGACAA..... 10264
576 erLeuIleGlyThrGlnTrpHisProGluAspLeuValaIleIleArg 592
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10264 ..... 10264
593 MetGluAspLysLysGluValIleProValThrHisAsnLeuThrLeuAr 609
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10265 .....CTGACACATCAAAATATTTGAA 10287
609 gLys.....ThrValThrGlyLeuAlaGlyAspArgThrLysAsp 623
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10288 ACCTTTGAAGTAACTTAATAAAAAAGCCGATTAATCAAGC..... 10330
623 heHisPheGluIleGluLeuLysAsnAsnLysGlnGluLeuLeuSerGln 639
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10331 .....CCACTTAAAGAGCGGAAATTCCTTTTAAACAGCA... 10363
640 ThrValLysThrAspLysThrAsnLeuGluPhe...LysAspGlyLysAl 655
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10364 .....CCAGATACGAGATTTGAATTTACCAAAAGATGGCAAGCA 10401
655 aThr.....IleAsnLeuLysHisGlyGln...SerLeuT 666
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

10402 AACGATACATTTGTTTGTGAAAACCTAAACACGAGGAATAATGTTCTAA 10451
666 hrleunginglyleuproglyglytyrserlyleuvalylsugluthasp 682
10452 CAGAAACCTTACCCAGAAAGATATCAGGGGTTAAAGAACCAATGAA 10501
683 .....serglyglytyrlylsvalylsvalnsersergingluval 696
10502 TTAATAATTCGTGAAAGATGGTTCAGTCACGATACATGCGGAAAAAGTAGC 10551
696 aasnalaThrValSer.....LysThrGlyIleThrSerAspGlu 710
10552 AGATGTTTATTTCTCGAGAGAAATATCAAAATTAATTACCTTAACGCTTA 10601
710 hrleualaphegluasnasnlysgluprovalValProthrGlyValAsp 726
10602 CG.....AACCAAGCAAAAGGTTCTTACCTGAACGCGGCGCATTA 10642
727 GlnlyslleasnlytyrleuAlaLeuile.....ValIlealagl 740
10643 GGACGCTTGCTGGTTTACTTGATAGCATTAATGATTCGTGATAGCGGG 10692
740 ylieSerleuglyle 745
10693 TGTATCTCTTTATTT 10708
seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A30874
seq_documentation_block:
ID A30874 standard; DNA: 2127 BP.
XX
AC A30874:
XX
XX 19-SEP-2000 (first entry)
XX
DE Streptococcus equi fibronectin binding protein, FNZ, coding sequence.
XX
KM Fibronectin binding protein; SFS; vaccine; horse; strangles; therapy;
XX equine upper respiratory tract disease; S. equi infection; FNZ; ds.
XX
OS Streptococcus equi.
XX
FH Key Location/Qualifiers
FT CDS 108..1901
FT /tag= a
FT /product= FNZ
XX
XX PN MO200037496-A1.
XX
XX PD 29-JUN-2000.
XX
XX PF 21-DEC-1999; 99MO-SE02448.
XX
XX PR 22-DEC-1998; 98SE-0004491.
XX
XX PA (GUSG/) GUSG B.
XX PA (LIND/) LINDMARK H.
XX PA (JACO/) JACOBSSON K.
XX PA (FRYK/) FRYKBERG L.
XX
XX PI Guss B, Lindmark H, Jacobsson K, Frykberg L;
XX
XX DR WPI: 2000-442641/38.
XX DR P-PSDB; Y90257.
XX
XX PT New protein useful for preparation of vaccines for treatment of
XX PT strangles caused by Streptococcus equi infection, is able to bind to
XX PT mammalian fibronectin.
XX
XX PS Disclosure: Page 17a-17b; 34pp; English.
XX
XX CC This sequence encodes the Streptococcus equi fibronectin binding
XX CC protein FNZ. The FNZ sequence was used to isolate the S. equi fibronectin

```

```

CC binding protein of the invention, designated SFS. SFS binds specifically
CC to mammalian fibronectin or its analogues or fragments. The protein, its
CC analogues or fragments may be used for the preparation of a vaccine that
CC protects horses against strangles (a world-wide distributed and serious
CC disease of the equine upper respiratory tract) caused by S. equi
CC infection. The antibody and/or antiserum may also be used for the
CC prophylactic or therapeutic treatment of S. equi infection in mammal,
CC especially horses. The use of vaccines containing the fibronectin binding
CC protein provides a more effective protection against S. equi infections,
CC with fewer side effects.
XX
XX SO Sequence 2127 BP; 718 A; 415 C; 466 G; 528 T; 0 other;

alignment_scores:
Quality: 180.00 Length: 716
Ratio: 0.562 Gaps: 33
Percent Similarity: 44.693 Percent Identity: 20.950

alignment_block:
US-09-494-297-2 x A30874 ..
Align seg 1/1 to: A30874 from: 1 to: 2127
68 SerGluThrArgTrpGlyGlySerGlyValArg...GlyHisPr 83
201 GCAGAGCAGCTTTATTTATGAGGTGGAATGATGAGACGACAAAGTTGCC 250
83 oTyTYrLYSGlnPheArgValAlaHisAspLeuArgValAsnLeuGlu 100
251 ATATTTT.....TTGACGATATGCGCTAAATAATGCTCCAAAGCGTGAT 294
100 lySerArgSerTYrGlnValTYrCysPheAsnLeuLYsValAAspPro 116
295 TAAAGACAGATGATGTGTTATGCTTTAAACAAATAATGATGTATGCCA 344
117 LeuGlySerAspSerSerValLYsLYsTrp..... 126
345 GATCAATGGGAAATCTATATACGCAATTTTAAATGATCAGATCTCCATA 394
127 .....TYrLYsLYSHisAspGlyIleSerThrLYsPheG 138
395 TAACGATTTACCTGATATGAGAAAAAAGTAGATATGATGCTATATTTA 444
138 lAspTYrAlaMetSerProArgIleThrGlyAspGluLeuAsnGlnLYs 154
445 AACAAATATGCTCCAGATTTACAAAAAGATATATGATATGCAAGTGT 494
155 LeuArgAlaValMetTYrAsnGlyHisProGlnAsnAlaAsnGlyIle 171
495 TTGGTGGCAGTTTAAAGTAAAGTACCCCACTAACAGTACACATATC 544
171 tGluGly.....LeuGluProLeuAsnAlaIleArgValThrGlnGlu 186
545 AACTAGCTACCATTTAAATAATGATCTCTAGAAAGATTACTCAATTAG 594
186 lValTYrTYrTYrSerAspAsnAlaProIleSerAsnProAspGluSer 202
595 CCATTGTGATTTTAGTAT.....ACTTTAACAAAAAGAAATAG 632
203 PheLYsArgGluSerGluSerAsnLeuValSerThrSerGlnLeuSer 219
633 CTTAAGATACGCGGTGCTTAATACTTA.....AACGATAT 667
219 uMetArgGlnAlaLeuLYsGlnLeuIleAspProAsnLeuAlaThrLYsM 236
668 GGAAGAAAAAGCTTTAGATTTTAAATCAGTAAAGAGAGCATCTAGC 717
236 eProLYsGlnValProAspAspPheGlnLeuSerIlePheGluSerGlu 252
718 TT...AAATCAGACAGATTAATTAATCATGATTTATTTATGTTATCAA 764
253 AspLYsGlnAspLYsTYrAsnLYsGlyTYrGlnAsnLeuLeuSerGly 269

```

```

765 AGTGGGGGATGACCATATGAAAGATTACCAAAATCTCTCGGCTTAC 814
766 yLeuValProThrlYsProThrlProGlyAspProPromeProPro 286
815 CTTAAATTCCTTAAAGAACCCCTAAAGCCT..... 842
286 snGlnProGlnThrlThrSerValLeuIleArgLysTyrAlaIleGlyAsp 302
843 .....CAGCTAGGTGCT 854
303 TyrSerLysLeuLeuGlnGlyAlaThrlLeuGlnLeuThrlGlyAsnVa 319
855 TTTAGT.....GGACATAATGG 871
319 IAsnSerPheGlnAlaArgValPheSerSerAsnAspIleGlyLysArg 336
872 AAATGGATTAAAGCGCCTTGAAGAGGATCATCGTTCCACAAAGAACTA 921
336 IeGlnLeuSerAspGlyThrlThrlThrlLeuThrlGlnLeuAsnSerProAla 352
922 ATGAA.....GATGTAAGAAAGACTT.....ATAGTTTCCATGGA 959
353 GlyTyrSerIleAlaGlu.....ProIleThr...PheLysVa 364
960 GGACCTCAGAGAGCGAGGCAACGAGATCCTTGGCAGGATGAAGCG 1009
364 IeGlnAlaGlyLysValThrlThrlIleAspGlyLysGlnIleGlnAsp 381
1010 TGAAGCTGGTGACCTGATACA.....CCTCAAAAGC 1041
381 roAsnLysGlnIleValGluProTyrSerValGlnAlaTyrAsnAspPhe 397
1042 CTAATGATCCATG.....CAGGCTCTGAAGCGGTAAAC..... 1076
398 GlnGlnPheSerValLeuThrlThrlGlnAsnLysValLysPheTyrAla 414
1077 .....TCTCCTATGTAAGAACAAACTATGTAGT.....AC 1108
414 alyAsnLysAsnGlySerSerGlnValValTyrCysPheAsnLysAsp 431
1109 CGAAGGATATCATGTCATCAATCAGCATCTTGAG.....GAAACCGAAG 1152
431 euLysSerProProAspSerGlnAspGlyLysThrlMetThrProAsp 447
1153 ATACTAACCCACCTGATATCATCTAGCGCGCTAGGAATGTGGA... 1199
448 PheThrlThrlGlnValLysTyrThrlHisIleAlaGly..... 460
1200 ...ACGATGAGATAGTACGAAACCCCTCATCTGATGGGATCGGCGCG 1246
461 .....ArgAspLeuPheLysTyrThrlValLysProArg..... 471
1247 TCTAGCTGGCGAATCAGAGAAACGACACCTAAACCGAGCAAAACGCGG 1296
472 .....AspThrAspProAspThrPheLeuLysHisIle 482
1297 GCGAAGGACCATGTCAGAGCAACAGAGATACA..... 1331
483 LysLysValIleGlnLysGlyTyrArgGlnLysGlnAlaIleGlnTyr 499
1332 .....CAAAAGGCATGCTGTGACAAATC 1354
499 rSerGlyLeuThrlGlnLysValAlaIleThrlGlnLeuAlaIleT 516
1355 TGTGGGACTATCGATCAGAAACACCAAAAGCCGAGGTGATGATG 1404
516 yTrPheThrAspSerAlaGlnLeuAspLysAspLysLeuLysAspTyr 532
1405 GTGGTCAGGCAACCATCGAGACAAACAGAGGACACACAAAA... 1448
533 HisGlyPheGlyAspMetAsnAspSerThrlLeuAlaValAlaLysIle 549

```

```

1449 ...GGCATGCTGTGACAAATCTGGCGGTACTATGAG..... 1481
549 uValGlnTyrAlaGlnAspSerAsnProProGlnLeuThrlAspLeuAsp 566
1482 .....TCAAGAGCACTTAAGAAACCTGAGCT..... 1508
566 hepHeIleProAsnAsnLysTyrGlnSerLeuIleGlyThrlGlnTyr 582
1509 .....ATGATTGGTGTGAG... 1523
583 HisProGlnAspLeuValAspIleIleArgMetGlnAspLysLysGlnVa 599
1524 .....GGCAAAAT 1531
599 ILeProValThrlHisAsnLeuThrlLeuArgLysThrlValThrlGlyLeu 616
1532 CATCGACTTCTCTGAAC...ACCCAATCAGGATATCTGGCGAGT 1575
616 IeGlnAspArgThrlLysAspPheHisPheGlnIleGlnLeuLysAsnAsn 632
1576 CTGGTGACACTAG.....GTAATTGAG..... 1598
633 LysGlnIleLeuLeuSerGlnThrlValLysThrlAspLysThrAsnLeuG 649
1599 .....GATACCAAGAAAGTCAAGATTAAT 1621
649 uPheLysAspGlyLysAlaThrlIleAsnLeuLysHisGlyLysSerLeu 666
1622 CATTTGGTGGGCAAGCAAAATCATGCACTTCTCGAGAGATACACGCCG 1671
666 hrlEuglnGlyLeuProGlnGlyTyrSerTyrLeuValLysGlnThrlAsp 682
1672 GTATGCTGTG...CAATCTGAGAGGCACCTACAATTTGCGAAGACACCAAG 1718
683 SerGlnGlyTyrLysValLysValAsnSerGlnGlnValAlaAsnAlaTh 699
1719 AAGCGCAACCTAAGCCTAAACCTGCACCTGCCCAATTTGTAAT..... 1763
699 rValSerLysThrlGlyIleThrSerAspGlnThrlLeuAlaPheGlnAsn 716
1764 .....GACGAA.....AAACCTA 1776
716 snLysGlnProValAlaProThrlGlyValAspGlnLys..... 728
1777 ACAAGGCACTCATCTCCACAGCAAGTATGAGCAACTCACCCCTA 1826
729 ...IleAsnGlyTyrLeuAlaIleValIleAlaIleAlaLysIleSerLeu 743
1827 AGCAATCATCGGTGCAATGTCAATGCTCTGTGCTATGTCGTCTTA 1874
seq_name: /cgnl_8/cgdata/geneseq/geneseq/NA1999.DAT:V82019
seq_documentation_block:
ID V82019 standard; DNA; 7650 BP.
XX V82019;
XX 21-JUN-1999 (first entry)
DE Moraxella catarhalis 4223 lfr region.
XX
XX lacioferrin receptor; lactoferrin binding protein; LBPl; LBp2;
XX lbpa gene; lbp gene; ORF3; infection; otitis media; sinusitis;
XX conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;
XX diagnosis; therapy; vaccine; Branhamella catarhalis; ss.
XX Moraxella catarhalis.
XX OS
XX Key Location/Qualifiers
XX -35_signal 64..69
XX FT /*tag= a
XX FT 98..103
XX FT -10_signal
XX FT /*tag= b

```





```

2502 AAAAATAAGCTAAGTTGATGCTGATTTTGACACCAACCTTAAGTGG 2551
614 YLEuAlagLYAsparGThrLysAspPheHisPheGluIleGluLeuYsa 631
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
2552 TAAATTAACATGAGAGAGGTGATATCGTCTTGTATATC.....AAA 2595
631 snAsnLYsGInGluLeuSerGInThrValLYsThrAspLYsThrAsn 647
      |||::|||::|||::|||::|||::|||::|||::|||::|||
2596 ATGGCAAAATGATGCGACAGGCTTTACCGCCAAAGCCGATGTCCAAAC 2645
648 LeuGluPheLYsAspGLYsAlaThrIleAsnLeuLYsHISGlyGluSe 664
      |||::|||::|||::|||::|||::|||::|||::|||::|||
2646 TATGTGTAAGAGTGGT.....AACACCAAGGTGC..... 2678
664 rLeuThrLeuGInGlyLeuProGInGlyTYrSerTYrLeuValLYsGluT 681
      |||::|||::|||::|||::|||::|||::|||::|||::|||
2679 .....GGTTTCTTATACACATCAAGATA 2703
681 hrAspSerGInGlyTYrLYsValLYsAlaSnSerGInGluValAla... 696
      |||::|||::|||::|||::|||::|||::|||::|||::|||
2704 TTGATGTCAAGGGCAATTTTGGCACAAATGGCGAAGAGTTGGCAGCG 2753
697 .....AsnAlaThrValSerLYsTh 703
2754 CAGTTACAGTACGACAAAGCGGATGCGATCATGACACCGCGAAAAAGC 2803
703 rGly 704
      |||
2804 AGCG 2807

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:V82021
seq_documentation_block:
ID      V82021 standard: DNA; 2718 BP.
XX
AC      V82021:
XX
DT      21-JUN-1999 (first entry)
XX
DE      Moraxella catarhalis VH19 lbp gene.
XX
KM      Lactoferrin receptor; lactoferrin binding protein; LBP1; LBP2;
KM      lbpA gene; lbpB gene; ORF3; infection; otitis media; sinusitis;
KM      conjunctivitis; pneumonia; bronchitis; tracheitis; emphysema;
KM      diagnosis; therapy; vaccine; Branhamella catarhalis; ss.
XX
OS      Moraxella catarhalis.
FH
FH      Key      Location/Qualifiers
FH      CDS      1..2718
FT      /tag=a
FT      /transl_except- (pos:847..849, aa:Thr)
FT      /transl_except- (pos:850..852, aa:Thr)
FT      /transl_except- (pos:853..855, aa:Thr)
FT      /transl_except- (pos:856..858, aa:Thr)
FT      /transl_except- (pos:859..861, aa:Thr)
FT      /transl_except- (pos:862..864, aa:Thr)
FT      /transl_except- (pos:868..870, aa:Thr)
FT      /transl_except- (pos:871..873, aa:Thr)
FT      /transl_except- (pos:874..876, aa:Thr)
FT      /transl_except- (pos:877..879, aa:Thr)
FT      /transl_except- (pos:880..882, aa:Thr)
FT      /transl_except- (pos:886..888, aa:Thr)
FT      /transl_except- (pos:891..891, aa:Thr)
FT      /transl_except- (pos:892..894, aa:Thr)
FT      /transl_except- (pos:895..897, aa:Thr)
XX
XX      W09855606-A2.
XX
XX      10-DEC-1998.

```

```

PF      02-JUN-1998: 98MO-CA00544.
XX
XX      08-MAY-1998: 98US-0074658.
PR      03-JUN-1997: 97US-0867941.
XX
XX      (CONN-) CONNAUGHT LAB LTD.
XX
XX      Du R, Klein MH, Loosmore SM, Wang Q, Yang Y;
PI      WPI: 1999-070266/06.
DR      P-SDB; W89421.
XX
PT      Lactoferrin receptor genes from Moraxella, especially M. catarhalis
PT      - useful to diagnose Moraxella infection e.g. to detect otitis media
PT      due to M. catarhalis infection and to immunise against such
PT      infections
XX
PS      Claim 8; Fig 16A-O: 202pp: English.
XX
XX      This polynucleotide comprises the lactoferrin binding protein lbpB
XX      gene of Moraxella catarhalis (Branhamella catarhalis) strain VH19.
XX      It encodes lactoferrin binding protein 2 (lbp2, see W89421). The
XX      CC genes and DNA sequences of the lactoferrin receptor (lfr) locus of
XX      CC Moraxella are useful for diagnosis, immunisation, and the
XX      CC generation of diagnostic and immunological reagents. Immunogenic
XX      CC compositions, including vaccines, based upon expressed recombinant
XX      CC lbp1 and/or lbp2 and/or ORF3, portions of these, or their
XX      CC analogues, can be prepared for prevention of diseases caused by
XX      CC Moraxella. M. catarhalis is a causative agent of otitis media and
XX      CC has been associated with sinusitis, conjunctivitis and inflammatory
XX      CC diseases of the lower respiratory tract, such as pneumonia, chronic
XX      CC bronchitis, tracheitis and emphysema.
XX
SQ      Sequence 2718 BP; 958 A; 585 C; 547 G; 628 T; 0 other;

```

```

alignment_scores:
      Quality: 176.00      Length: 796
      Ratio: 0.515      Gaps: 40
Percent Similarity: 42.965      Percent Identity: 20.603

```

alignment\_block:

US-09-494-297-2 x V82021 ..

Align seg 1/1 to: V82021 from: 1 to: 2718

```

84 TYrTYrLYsGInPheArgValAlaHisAspLeuArgValAsnLeuGlu 100
      |||::|||::|||::|||::|||::|||::|||::|||::|||
493 TATTTGATTAATTCGCCAAATATCCGATCTG.....CACCTAGAA 536
100 ySerArgSerTYrGInValTYrCYsPheAsnLeuLYsLYs..... 114
      |||::|||::|||::|||::|||::|||::|||::|||::|||
537 CAGCGAG.....CATGTGTTGATGCTTAATAAACCAATATA 574
115 .....PheProLeuGlySerAspSerValLYs..... 125
      |||::|||::|||::|||::|||::|||::|||::|||::|||
575 TCAAAATATATGCTTATGCTGATGTCATCATCTGCCAAACCAAC 624
126 .....TYrTYrLYsHISAspGlyIleSerThrLYs.....PheG 138
      |||::|||::|||::|||::|||::|||::|||::|||::|||
625 TACATGAAATTATCAACAGAACAAACATCAAAAACCAAGCGGA 674
138 uAspTYrAla.....MetSerProArgIleThr.... 147
      |||::|||::|||::|||::|||::|||::|||::|||::|||
675 TGATTTATCAAAACATTCGTTTGGCTATATGAGCTAAGACGTGGAGC 724
148 .....GlyAspGluLeuAsnGInLYsLeuArgAlaValMet 159
      |||::|||::|||::|||::|||::|||::|||::|||::|||
725 TAAATTAATAAAAGGTGACAGACCCAGAGCGCAAGAACCCGTCAT 774
160 TYrAsnGly.....HisProGInAsnAlaAsnGlyIleMe 171
      |||::|||::|||::|||::|||::|||::|||::|||::|||
775 TTCACACACCTACTTTATTTATCATGCTGAGATGCCACACCATCT 824

```

```

171 tgluglyleuGluproLeuasnalaileargValThrGlnGluAlaValT 188
      :      :      :      :      :      :      :      :
825 GCCA.....AAGCGGGTAATTTGACTATGAGGCGCAATT 859
188 rpyrTyrserspAsnAlaPro.....IleserAsnProasp 200
      :      :      :      :      :      :      :      :
860 GGTGTATCTGCACCGATGTCAAAAACGCCCATTTTATGATTAACAGAC 909
201 Glu.....SerPheLysArgLusSerGlnSerLys 210
      :      :      :      :      :      :      :      :
910 GATAAATAGGCACTTATTTTAACTCAACGAAAAATCAATGAGGCGA 959
210 nleuValSerThrSerGlnLeuSerLeu..... 219
      :      :      :      :      :      :      :      :
960 TTTCGTAGTCACACACATTTATCTAAACAGCTTAAATATAACACA 1009
220 .....MetArgGlnAla 223
1010 CCCCAGCCACTTATAGCGTGGACTTTGATCAAAATACCTTAAAGCGCAA 1059
224 leuLysGlnLeuIleaspProasnLeuAlaThrLysMetProLysGlnVa 240
      :      :      :      :      :      :      :      :
1060 TTGTCTATTATGACAAACCAACAGCAACAGCCGATGGCGTTATAT 1109
240 LProaspAspPhe.....GlnLeuSerIlePheG 250
      :      :      :      :      :      :      :      :
1110 CAGAAAGTCAGTTGATACCGCAAAAAAGTCATGAAGCCGATGTCATG 1159
250 luserGluaspLys.....GlyaspLysTyrsnLysGlyTyrgLnaSn 264
      :      :      :      :      :      :      :      :
1160 AGATTGACGCCAAGATTAAATGCAACCGCTTACTGACACGCCCAATCT 1209
265 leuLeuSerGlyGlyLeu.....ValProth 273
      :      :      :      :      :      :      :      :
1210 TTGATTATGATTAACCAACCAATACCGCACTTTTGTAAAGAGCTGCTC 1259
273 rLysProProThrProGlyaspProPheProProasnGlnProGlnT 290
      :      :      :      :      :      :      :      :
1260 CAATAAAGCCATCCCAACAACCA.....GACCCCAAC..... 1293
290 hrThrSerValLeuIleArgLysTyraIleGlyaspTyrsLysLeu 306
1294 .....TCGATATACG 1302
307 leuGlnGlyAlaThrLeuGlnLeuThrGlyaspAsnValasnSerPheG 1323
      :      :      :      :      :      :      :      :
1303 CTAAAGAGCGGGTTTATGCTGAGCTGGCGATGAGCTG..... 1341
323 nAlaArgValPheSerSerAsnAspIle.....G 333
      :      :      :      :      :      :      :      :
1342 GCGGTAAATTTTATCCATGACAAACGCAACTTTGTGCTTTGTGTG 1390
333 LysGlnArgIleGlnLeuSeraspGlyThrThrLeuThrGlnLeuasn 349
      :      :      :      :      :      :      :      :
1391 GCACACGACGAAAAACGACCACTGTGCGCAAAAAACGCTATTTT 1440
350 SerProAlaGlyTyrserIleAlaGluProIleThrPheLysValGluAl 366
      :      :      :      :      :      :      :      :
1441 AGT...ACAGATTGAA.....AAACCCAGCACCAAGCTTTGTTGCCAA 1481
366 agLysValValTyrrThrIleIleaspGlyLys.....GlnIleg 379
      :      :      :      :      :      :      :      :
1482 TGAAGAGATTGGTACATTTATGACGTAATAAAGTTAATGATGATGCA 1531
379 lAsnProasnLysGlnIleValGluProTyrsValGluAlaTyrsn 395
      :      :      :      :      :      :      :      :
1532 ATAAATCAAAATGAAGATGAACGTCTCCCTGCACTAATAAAGATATTAT 1581
396 AspPhe.....GluGluPheSerVa 402
      :      :      :      :      :      :      :      :
1582 GAATATATATTATGACGACCAACCAAAACATTTCCCAAAAAAATAAAGC 1631
402 lLeuThrThrGlnAsnTyraIa.....LysPheTyrr 413
      :      :      :      :      :      :      :      :
1632 CAGCGTCCAAAAAACCCCTGCTATTTTGGTCAGCATGATAGTTTAT 1681
413 yr.....AlaLys.....AsnLysAsnGly 419
      :      :      :      :      :      :      :      :
1682 TTAATGTAACATATTATGACTTATACGCCAAAGACAAACAGCTGGT 1731
420 SerSerGlnValValTyrcysPheasnAlaaspLeuLysSerProAs 436
      :      :      :      :      :      :      :      :
1732 GTCTCCCAAGATTAACACACCAATTAAGATATTTGGCTAAATACCAGA 1781
436 pSerGluaspGlyGlyLysThrMetThrProaspPheThr..... 450
      :      :      :      :      :      :      :      :
1782 TGCC.....AAAGTAAACAGACAAATAAAGTTACCAAAATCG 1819
450 ..... 450
1820 TTCTACAAACAGCCAAAGATTAAGCGGTATACCGCATTCATGCCAAAGC 1869
451 .....GlyGluValLysTyrrThrHisIleAlaGlyAr 461
      :      :      :      :      :      :      :      :
1870 TATGACCAATCACTTTGGTGAAGTATTTGATATATGATTAACAAGCAA 1919
461 gaSPLeuPheLysTyrrThrValLysProArGaSPThrAspProaspThr 478
      :      :      :      :      :      :      :      :
1920 CCCAAACAGAGTATTGTCGAAGCGGTCAAGCGATGTCAGTACTC 1969
478 heLeuLys.....His 481
1970 AGCTGCCAGTCAGGTAAATTCACCTATATAGTCTTTGGCAGCTAC 2019
482 lLeuLysValIleGlnLysGlyTyrrArgGlnLysGlyGlnAlaIleg 498
      :      :      :      :      :      :      :      :
2020 CTGACCCGAAAAAAGACAAAGTTATAGCAAAAGTACAGATACATCA 2069
498 uTyrsGlyLeuThrGlnThrGlnLeuArgAlaAlaThrGlnLeuAla 515
      :      :      :      :      :      :      :      :
2070 GCAAAAAGCTTTAAAGAT..... 2088
515 lEtyrTyrrPheThr..... 519
2089..TATATATTGACCAAGACTTATCCACAGATGACGATGATGAC 2136
520 AspSerAlaGlnLeuaspLysaspLysLeuLysasp...TyrrHisGlyP 535
      :      :      :      :      :      :      :      :
2137 GATAGTTTGACCGCATCTGATGATTCACAAGATGATATACATGCGCA 2186
535 eGlyaspMetAsnaspSerThrLeuAlaValAlaLysIleLeuValGluT 552
      :      :      :      :      :      :      :      :
2187 TGATGATTTG.....ATTGCATCTGATG 2209
552 yrAlaGlnaspSerAsnProProGlnLeuThrAspLeuaspPheIle 568
      :      :      :      :      :      :      :      :
2210 ATTACAGAGATGACACAGATGGCGATGACGATTAAGATGATTG... 2256
569 ProasnAsnLysTyrglnSerLeuIleGlyThrGlnIrrPheProG 585
      :      :      :      :      :      :      :      :
2257...GGTATGGTGACATGATGACGCGCGCAAGCAATGATATCATGAGG 2303
585 uAspLeuValaspIleIleArgMetGluaspLysLysGluValIleProV 602
      :      :      :      :      :      :      :      :
2304 TAAT.....ATTGCGCCCTGAATTTGAAACAAATACTTGCCCA 2341
602 al.....ThrHisasnLeuThrLeu..... 608
2342 TTAATGACCTACTCATGAATAAACCTTTCCTAGATGCTAAATAATAG 2391
609 .....ArgLysThrValThrGlyLeuAlaG 617
2392 GCTAAGTTTGATTAACCTTGAACACCAACAGCGCTAAGTAAATATAA 2441
617 yaSPArgThrLysaspPheHisPheGlnIleGlnLeuLysasnLysG 634

```



```

        |||::: |::: |||||:::|||||
    2442 CGATGAGAGGTGATTACCGTCTTTGCATATC.....AAAATGGCAAA 2485
                                ::::|
634 IngluleuLeuSerGIInthrValIysThrAspLysThrasnLeuGlutpe 650
      ::: |::: |||||:::||| :::
2486 TTGATGCCACAGCATTTACC GCCAAGA CCGATGTC CAACA TATCGTGA 2535
                               :|
651 LysAspGlyLysAla Thr Ileasn Leu ysh Sgi Glu Ser leu Thrle 667
      ::| |::: ||| ::: ||
2536 GAAGTG GGT .....AAC AACCA AGGTG GC..... 2593
                        |||::: ||| :::||::: |||
667 uGIInglyLeuProGIugLyTrSe rTyLeuVallYsglIn thr Aspserg 684
                          .GGTTCTTA TCACACTCAA AAgAtTAG ATGATT TA 2593
2594 AC GGCCAA TTTTT TGCCAC AAATA TGCCGA AGATTTG CA 2631
                                     ::|||
seq_name: /cgnl_8/gcgdata/geneseq/genesegN/A1999.DAT.X20201
seq_documentation_block:
ID   X20201 standard; DNA; 4249 BP.
XX
XX     X20201;
XX
XX     20-APR-1999 (first entry)
XX
XX     Enterococcus faecalis EF104 gene fragment.
DE
XX
XX     Enterococcus faecalIs; infection: vaccine; immune response; diagnosis;
KM detection; attenuation; antigenic; ss.
XX
XX     Enterococcus faeacals.
OS
XX
XX     M09850554-AZ.
PN
XX
XX     12-NOV-1998.
PD
XX
XX     04-MAY-1998; 98MO-USO8959.
PF
XX
XX     14-NOV-1997; 97US-0066009.
PR     06-MAY-1997; 97US-0044031.
PP     16-MAY-1997; 97US-0046655.
PX
XX
XX     (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX     Bailey C, Choi GH, Hiromocky J A, Kunsch CA;
PI
XX
XX     WPI; 1999-070095/06.
DR
XX
XX     P-PSDB; Y00211.
DT
XX
XX     New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
TR for prevention or attenuation of Enterococcus infection
TS
XX
XX     Claim 1; Page 202-203; 301pp; English.
TX
XX
XX     The present sequence encodes an antigenic polypeptide fragment
CC isolated from Enterococcus faecalis. The present invention describes
CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
CC The proteins can be used in vaccines for preventing or attenuating an
CC infection caused by a member of the Enterococcus genus in an animal.
CC They can also be used for detecting Enterococcus antibodies in a sample
CC acids. Products from the present invention can be used for detecting nucleic
CC screening compounds to identify agonists and antagonists of E. faecalis
CC protein activity.
CX
XX
XX     Sequence 4249 BP; 1579 A; 779 C; 815 G; 1069 T; 7 other;
SO
alignment_scores:
```

```

      Quality:   161.00      Length:   755
      Ratio:     0.434       Gaps:     40
      Percent Similarity:   49.139      Percent Identity: 21.060

alignment_block:
US-09-494-297-2 x X20201      ..

```

Align seg 1/1 to: X20201 from: 1 to: 4249

[illegible]



PA (HUMA-) HUMAN GENOME SCI INC.

XX Bailey C, Choi GH, Hromockyj A, Kunsch CA;

PI WPI: 1999-070095/06.

DR P-PSDB; Y00210.

XX New isolated Enterococcus faecalis polynucleotides - used to develop  
PT products for the detection of Enterococcus and for use in vaccines  
PT for prevention or attenuation of Enterococcus infection

PS Claim 1; Page 200-201; 301pp; English.

XX The present sequence represents a gene isolated from  
CC Enterococcus faecalis. The present invention describes genes, proteins  
CC and antigenic polypeptides isolated from E. faecalis. The proteins can  
CC be used in vaccines for preventing or attenuating an infection caused  
CC by a member of the Enterococcus genus in an animal. They can also be  
CC used for detecting Enterococcus antibodies in a sample. The nucleotide  
CC sequences can be used for detecting Enterococcus nucleic acids.  
CC Products from the present invention can also be used for screening  
CC compounds to identify agonists and antagonists of E. faecalis protein  
CC activity.

XX Sequence 4359 BP; 1614 A; 794 C; 838 G; 1106 T; 7 other;

# alignment\_scores:

Quality: 161.00 Length: 755  
Ratio: 0.434 Gaps: 40  
Percent Similarity: 49.139 Percent Identity: 21.060

# alignment\_block:

US-09-494-297-2 x X20200 ..

Align seg 1/1 to: X20200 from: 1 to: 4359

```

39 AAlaLeuValThrSerMetValGlyAlaLysThrValPheGlyLeuValG1 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1546 TCATTAAGTACACCTGTAATGGTCTCAATTAAGCAATTCATTAAGTATC 1595
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 user..... SerThrProAsnAlaIleAsnProAspSerSerSerg 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1596 CGATCAATTAATTAACCAATTAAGTGTGTTAACTCT...TTGAATGCTG 1642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
69 luytArGtrPrTyGlyTyrGlnSerTyValAlaGlnHisProTyLys 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1643 AAATGCTGGGTAATTATGATCA..... 1668
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 LysGlnPheArgValAlaHisAspLeuArg.....ValAsnLeuGluG1 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1669 .....AATGGCTGCTATTCATCAAGAACACTGCTCAGTTAATGGG 1709
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
100 ySerArgSerTyGlnValTyrCysPheAsnLeuLysLysAlaPhePro. 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1710 AAGCAAGAGAACCCGATTCAAATTTAGAAATTAAGTAAGCAATCCTA 1759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
117 .....LeuGlySerAspSerValLysLysTyrPheLysLys 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1760 ATTAATCTTCTTACGAGCTACAAAGAAATTTATTTTATTACAAAG... 1806
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 HisAspGlyIleSerThrLysPheGluAspTyrAlaIleMetSerProArg1 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1807 .....TTAGCAAGC.....GATTATACAGTAACGCAACGCTC 1838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 eThnGlyAspGluLeuAsnGlnLysLeuArgAlaValMetLysnGln 163
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1839 AGATGGTTCACCTTATTAAGTTCACCTACGCCAATAACCAAGAAATCCAA 1888
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 lSProGlnAsnAlaAsnGlyIleMetGlnGlyLeuGlu..... 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1889 TTCCAATTTGTTTAAATTAATGTGCCACAGTACTTTGCCAAAGATAAAGT 1938

```

```

176 ...ProLeuAsnAlaIleArgValThrGlnGluAlaValTrpTyTyrS 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1939 ATCCAGTCCGATACGATACCGATACATGATGCTTAAAGTTTAATCC 1988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 rAspAsnAlaProIleSerAsnProAspGluSerPheLysArgLysSerg 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1989 AGTGCATACGACAGTACTACT.....AATAGTAAGGCTGCTCTG 2029
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
208 luserAsnLeuValSerThrSergLuserLeuMetAlaGlnAlaLeu 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2030 AACGACACTTCAAAGTGAATAAAATCAATTCCTTGCATGACGAAAT 2079
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 LysGlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnValPr 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2080 GATTCCTTTGACTCAGTACAGGCTGTCACAAATTCAGTGGCCGC... 2127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 oAspAspPheGlnLeuSerIlePheGlnSerGluAspLysGlySplyst 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2128 ...GATGTCCTTTTGACATTTATGATGTTCAAC.....GATCAGG 2167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 yRAsnLysGlyTyrGlnAsnLeuLeuSergLysGlyLeuValProThrLys 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2168 TAGATTCATTTATTCACAAATCTGGGACCGCGTCAATCTTGTGATAA 2217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 ProProThrProGlyAspProPrometPro.....ProAsnGlnProG1 289
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2218 CCAATGAGCGCCAAACAGCCCTGATATCCAAACGATTACTTTGACGAAA 2267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 nThrThrSerValLeuIle.....ArgLysTyrAlaI 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2268 TACCAATAGATTACACGTTTGTGTTTGGAAAACCAACAAAGTTACATTA 2317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 lGlyAspTyrSerLysLeu.....LeuGlnGlyAlaThrLeuGln 313
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2318 TT...GAGTATATAAACGCCAATGCTGATCGACGTGCCACTCTTAT 2364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 LeuThrGly.....AspAsnValAsnSerPheGlnAl 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2365 ATTAACAGGACAGCGAAGAAACCAACATCGAATTAATTAAGGCTCTGCG 2414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 aArgValPheSerSerAsnAspIleGlyGluArgIleGluLeuSerAspG 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2415 TTCGGTTTCCTTCAAAAT.....GAGCGTTAGACATTTTGAGTG 2455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
341 lYthrTyThrLeuThrGlnLeuAsnSerProAlaGlyTyrSerIleAla 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2456 CAACA.....CAAGCGCGG 2469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358 GluProIleThrPheLysValGluAlaGlyLysValTyrThr...IleI 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2470 AATCCAAACATTAATAAAGTAAACAAAACGACAGTAAACAAAAAATAT 2519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
373 eAspGlyLys.....GlnIleGluAsnProAsnLysGlnIleValGluP 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2520 TGATTAATAAACACATGCTGTGAAAATCAACGATTAATTAACACCAA 2569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388 rOtyrSerValGluAlaTyrAsnAspPheGlnGluPheSerValLeuThr 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2570 AAGGACAAACCAATGCTCAATCGATTTGAATTCATATACCGTG... 2613
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
405 ThrGlnSerTyAlaLysPheTyTyAlaLysAsnLysnGlySerSe 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2614 .....AAAGCGTGGCC 2624
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 rGlnValValTyrCysPheAsnAlaAspLeuLysSerProProAspSerg 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2625 AGAAGATGCTTAT.....TCATTAAGAAGACTA 2653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
438 lAspGlyGlyLysThrMetThrProAspPheThrThrGlyGlu..... 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2654 CAAACGTCGCAAGTCAATTTTAAAGACTATACATTCACGAAACATAT 2703
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 ...ValLysTyThrHisIleAlaGlyArgAspLeuPheLysTyThrVa 468

```



55 user.....SerThrProAsnAlaIleAsnProAspSerSerSerg 69  
 6984 CCAATCAATATATTGAACCAATATGCTGTATATCT...TTGAATGCTG 9730  
 69 lutyAlaGlrPyrGlyItyrGlyuSerItyrValaIrgLysHisProItyrItyr 85  
 9731 AAAGCTCTGGGTAAATTAATGATCA..... 9756  
 86 LysGlnPheArgValaIAlaHisAspLeuArg.....ValAsnLeuGluG1 100  
 9757 .....AATGGTGCCATTCATCAACAGAACAACTGCTCTCATATGGG 9797  
 100 ySerArgSerItyrGlnValItyrCysPheAsnLeuLysItyrAlaPhePro. 116  
 9798 AACCAAGAGAAACCGAATTCAAATTTAGAAATTAAGTAAGCATCTCA 9847  
 117 .....LeuGlySerAspSerSerValLysItyrPyrItyrLys 129  
 9848 AATTATCTTCATTACGATTCACCAAGAAATTTATTTATTAACAAG... 9894  
 130 HisAspGlyIleSerThrLysPheGluAspItyrAlaMetSerProArgI1 146  
 9895 .....TTAGAACG.....GATTATACAGTACGCCCAACGC 9926  
 146 eThrGlyAspGluLeuAsnGlnLysLeuArgAlaValMetItyrAsnGlyH 163  
 9927 ACATGGTTACGTTAATTAAGTTCATACGCCAATTAACCAAGAAATCCAA 9976  
 163 lserProGlnAsnAlaAsnGlyIleMetGluGlyLeuGlu..... 175  
 9977 TTCCAATGGTTTAAATTAATGTCAGATAGTTCGCCAAAGATAAAAGT 10026  
 176 ...ProLeuAsnAlaIleArgValThrGlnGlnAlaValItyrPtyrTyrSe 191  
 10027 ATCCCACTGCATACATACCATCCGATTAACATGAGTCTGGAAGTTTAATCC 10076  
 191 rAspAsnAlaProIleSerAsnProAspGluSerPheLysArgGluSerG 208  
 10077 AGTGTATACGACAGTAATCTACT.....AATAGTAAGCGTGGTCTG 10117  
 208 luserAsnLeuValSerThrSerGlnLeuSerLeuMetArgGlnAlaLeu 224  
 10118 AACGAACACTTAAAGTAGTAAATAAATCAATTCCTGTCAATGCACGAAT 10167  
 225 LysGlnLeuIleAspProAsnLeuAlaThrLysMetProLysGlnValPr 241  
 10168 GATTCTTTGACTCATAAGCGTCCGTACAAAATTCACAGCTGGCGCC... 10215  
 241 oAspAspPheGlnLeuSerIlePheGluSerGluAspLysGlyAspLysT 258  
 10216 ...GATGTTCTTTTGGACATTTATGATGTTTCAAAC.....GATCAGG 10255  
 258 yTrAsnLysGlyItyrGlnAsnLeuLeuSerGlyGlyLeuValProThrLys 274  
 10256 TAGATTCAATTTATCCACAAATACTGGGACCGGTCATATCTTGATAA 10305  
 275 ProProThrProGlyAspProPheMetPro.....ProAsnGlnProG1 289  
 10306 CCNAATGACCCCAACAGCCCTGATATCCACGATTACTTTTGACAGAAA 10355  
 289 nThrThrSerValLeuIle.....ArgLysItyrAlaI 300  
 10356 TACCAATATTAACGTTTGATTTGGAAAAACCAACAAAGCTTACATTA 10405  
 300 leGlyAspItyrSerLysLeu.....LeuGluGlyAlaThrLeuGln 313  
 10406 TT...GAGATATAAAAACCGCAATGCTGATGACGATGCCAATCTTTAT 10452  
 314 leuThrGly.....AspAsnValAsnSerPheGlnAl 324  
 10453 ATTAACAGGACACGCAACCAACCAATGATTAATATGAAGGCTCTGC 10502  
 324 aArgValPheSerSerAsnAspIleGlyGluArgIleGluLeuSerAspG 341  
 10503 TTCGGTTTCTGTTCAAAAT.....GAACGGTTAGCAATTTGACGTG 10543  
 341 lYThrItyrThrLeuThrGluLeuAsnSerProAlaGlyItyrSerIleAla 357  
 10544 CAAACA.....CAACGGCG 10557  
 358 GluProIleThrPheLysValGluAlaGlyLysValItyrThr...IleI1 373  
 10558 AATCCAAACATTAATAAATGTATACAAAACAGACAGTAAACAAATAAT 10607  
 373 eAspGlyLys.....GlnIleGluAsnProAsnLysGluIleValGluP 388  
 10608 TGATATATAAACACATCGGTGTAATAAATCCAAACGATTAATTAACACAA 10657  
 388 roItyrSerValGluAlaItyrAsnAspPheGluGluPheSerValLeuThr 404  
 10658 AAGGCACAAACCATGCTCAAAATCGATTGAAATTCATTACCGTG... 10701  
 405 ThrGlnAsnItyrAlaLysPheItyrItyrAlaLysAsnLysAsnGlySerSe 421  
 10702 .....AAGGGGTGCG 10712  
 421 rGlnValValItyrCysPheAsnAlaAspLeuLysSerProProAspSerg 438  
 10713 AGAAGATGCTTAT.....TCATTAGAGAAAGACTA 10741  
 438 lAspGlyGlyLysThrMetThrProAspPheThrItyrGlu..... 452  
 10742 CAACGGTCCGAAGCATATTTAAAGACTATATACATTCACAGAAACATT 10791  
 453 ...ValLysItyrThrHisIleAlaGluArgAspLeuPheLysItyrItyrA 468  
 10792 AGATTGATTAACAATACGCTCTCTGCAAAACGGTGGCCAAATCTATACAA 10841  
 468 lLysProArgAspThrAspProAspThrPheLeuLysHisIleLysLysV 485  
 10842 AACCAACATCATGACTCGAACAATTGAAACCAAGATGCTGTACAGAAAA 10891  
 485 alIle.....GluLys 488  
 10892 AAGTCAACCTCGGCCCAATACATTGAATTTCTCAGAGGTGATGGGAA 10941  
 489 GlyItyrArgGluLysGlyGlnAlaIleGlyItyrSerGlyLeuThrGluTh 505  
 10942 GGTATGTTTATTTAGCACTGCGACATTCCTACAGCATTAACAGTAGAGA 10991  
 505 rGlnLeuArgAlaAlaThrGlnLeuAlaIleItyrItyrPheThrAsp... 520  
 10992 TGAAAAACAAGCAATTTGCGAAGGTTCTTTTAACTAATGTAATGTCA 11041  
 521 .....SerIleGluLeuAspLysAspLysLeuLysAspItyrHisGly 534  
 11042 CGCATACACAAACCGAATTTTACACAGAGGAAAGAGTAATATCTCTTT 11091  
 535 PheGlyAspMet...AsnAspSerThrLeuAlaValAlaLysIleLeuVal 550  
 11092 GATGCCATATACAGAGTATATATCTTTGCGAGTACCAATGTACCCGA 11141  
 550 lGluItyrAlaGluAspSerAsnProProGlnLeuThr..... 562  
 11142 GGAATATTCGCGATGAGAGAG.....TATTTGACAGGAAAGGCATTA 11188  
 563 .....AspLeuAspPhePheIlePro..... 569  
 11186 AGCTGGTCAAGAGAGACCAACCACTAAATAATTCATAAGAAACAATT 11235  
 570 AsnAsnAsnLysItyrGln.....SerLeuIleGlyThrG1 581  
 11236 GATCACAGCTGTTTACAAAGTCAAAATTCACAGATTATATGCGCGATTC 11285  
 581 nThrHisProGluAspLeuValAspIleIleArgMetGluAspLysLys. 597

```

11286 ATGGAACCAAGAG.....AACTTTGTTTCAGCAACGATTAACAG 11329
598 ..GluValIleProValThrHisAsnLeuThrLeuArgLysThrValThr 613
11330 GTCAAGACGTCCTTCGAA.....AAAATCACTGTTTCA 11364
614 GLLLeuAlaGlyAspArgThrLysAspPheHisPheGluIleGluLeu 630
11365 GGCCAATGTTATATAACACAAAGCGGTTTATTCACATTTTTCACGTGA 11414
630 sAsnAsnLysGlnGluLeuLeuSerGlnThrValLysThrAspLysThra 647
11415 CGAAGTAAGAAAGAAACAGCCTATGTGACCGTCAACCCGACCAATCTA 11464
647 snLeuGluPheLysAspGlyLysAlaThrIleAsnLeuLysHisGlyGlu 663
11465 AGTTAGAGGTCAAGAT.....ACAACGATTATATGTT.....GGTGAAT 11502
664 SerLeuThrLeuGlnGlyLeuProGluGlyTyrSerTyrLeuValLysG1 680
11503 TCGTGGAAA.....CCAGAA.....GATAATTTTCGTTTCAGC 11534
680 wThrAspSerGluGlyTyrLysVal.....LysValAsnSerGlnG 694
11535 GACGACAAACAAACAGCTACAGCGCTTTCGAAAAAATGATGTTTCAGC 11584
694 LuValAlaAsnAlaThrValSerLysThrGlyLThrSerAspGluThr 710
11585 GAACAGTGAAAT.....GTTCATTAATAATAGCC.....GATTATGAA 11619
711 LeuAlaPheGluAsn 715
11620 ATTGCTATATAAAAT 11634
seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT:T68844
seq_documentation_block:
ID T68844 standard; DNA; 5547 BP.
AC T68844;
XX
XX
XX 29-JAN-1998 (first entry)
DE Photorhabdus luminescens insect toxin TcdA11.
XX
XX Insecticide; insect; toxin; pest control; biological control;
XX Photorhabdus luminescens; TcdA; Southern corn rootworm;
XX Colorado potato beetle; Western corn rootworm; meal worm;
XX boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;
XX cabbage looper; codling moth; corn earworm; European corn borer;
XX tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;
XX Diptera; Dictyoptera; Acarina; Homoptera; ds.
XX
XX Photorhabdus luminescens strain W-14 (ATCC 55397).
OS
XX
XX WO9717432-A1.
XX
XX 15-MAY-1997.
XX
XX 06-NOV-1996; 96WO-US18003.
XX
XX 28-AUG-1996; 96US-0705484.
XX 06-NOV-1995; 95US-0007255.
XX 28-FEB-1996; 96US-0608423.
XX
XX (MISC ) WISCONSIN ALUMNI RES FOUND.
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fatig R;
XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL,
XX Petrelli J, Roberts JL, Rochelleau TA, Schoonover S;
XX Strickland JA;
XX
XX WPI; 1997-281022/25.

```

```

DR P-PSDB; W17900.
XX
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them -
XX can be genetically engineered into insect larvae food and plants for
XX insect control
XX
XX Claim 38; Page 201-208; 276pp; English.
XX
XX This genomic DNA sequence encodes the 209.2 kDa TcdA11 insecticidal
XX toxin protein (see W17899) of Photorhabdus luminescens W-14.
XX It is derived from a full-length clone (see T68844) obtained from a
XX cosmid library. A 282.9 kDa protein (see W17899) is proteolytically
XX processed into TcdA11 and TcdA11 (see W18301) components.
XX Claimed toxins of P. luminescens (see W17871, W17884-89, W17899-900,
XX W18301-06) can be produced by recombinant DNA methods and applied to,
XX or genetically engineered into, insect larvae food and plants for
XX insect control. The Photorhabdus toxins are particularly effective
XX against Southern corn rootworm, Colorado potato beetle, Western corn
XX rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet
XX armyworm, black cutworm, cabbage looper, codling moth, corn earworm,
XX European corn borer, tobacco hornworm and tobacco budworm
XX (Lepidoptera) and are also active against insects of the orders
XX Hymenoptera, Diptera Dictyoptera, Acarina and Homoptera. (All
XX claimed).
XX
XX Sequence 5547 BP; 1710 A; 1219 C; 1115 G; 1503 T; 0 other;
SQ
alignment_scores:
Quality: 154.00 Length: 809
Ratio: 0.424 Gaps: 37
Percent Similarity: 44.870 Percent Identity: 18.294
alignment_block:
US-09-494-297-2 x T68844 ...
Align seg 1/1 to: T68844 from: 1 to: 5547
157 AlavAlMeTYrAsnGLYHISProGlnAsnAlaAsnGLYIleMeTGluG1 173
||| :||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
2386 GCACACAATTTGAATGTCGCCGCCACAGCGCGTTCGCCCTTGTC...GG 2432
173 YLeuGluProLeuAsnAlaIleArgValThrGlnGlu...AlavAlTrPT 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2433 GCTGATTAATATTCATCAATCAAGAAAGACACGCGCTATGCCAGTGGC 2482
189 yTYrSerASP.....AsnAlaProIleSerAsn 198
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2483 AAAACGGCGAGCGGTATTAACCGCGCGGTGAATTCACAACAGCGTAAAT 2532
199 ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr.. 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2533 ACATTACACGCTTTTCGATGAATTCGCGAGTCGCCGATTAACACACTTA 2582
215 .....SerGluLeuSerLeuMetArgGlnAlaLeuLysGln..... 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2583 CTATATCCGTCACACTCGCAAGCAGCGCGGTATTAAGAAACCGTGATG 2632
227 .....LeuIleAspProAsnLeuAla..... 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2633 ACTTGATCAATACTTACTGATGATATCATCAGTTTTCGCGGCACATAAAA 2682
234 ...ThrLysMetProLysGlnValProAspAspPheGlnLeu..... 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2683 ACCACCGGATCGCCGAAGCATTTGCC...AGTATTCAACTGTATGCTCA 2729
247 .....SerIlePheGluSerGluAspLysGly..... 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2730 CCGGGCATTTGGAAATGTGAGAAATGCCAATTCGGGGTATATCAGCC 2779
256 .....AspLysTyrAsnLysGlyTyrGlnAsnLeu 265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2780 GCCAATTCCTTATGACTGGGACAAATACATAAACCCTACAGCACTTGG 2829

```

[illegible]

```

4558 CTAATTGCCCGCAGTTGGTTGCACGCCGCCACC.....ACCGGAAT 4598
705 ehtsreraspgluhtleualaphelulsnanlysgluptovalalp 722
4599 CGAATCAATTCGTGATGGAACCTCAAGATATTCAGAACCC..... 4641
722 rothrglyalaspgluhtleualaphelulsnanlysgluptovalalp 738
4642 .....CAGTTAGCAAGGTTCTATGCTAGCTGATGATA 4677
739 AlaGlyIleSerLeuGlyIleTptGly 747
4678 CCTCCCTATACCTATCACTCATGCT 4704

seq_name: /cgnl_8/gcdata/geneseq/geneseqn/NA1998.DAT:V29930
seq_documentation_block:
ID V29930 standard; DNA: 5547 BP.
AC V29930:
XX 07-AUG-1998 (first entry)
XX tcd gene from the tcd genomic region encoding protein TcdAll.
DE tcd gene from the tcd genomic region encoding protein TcdAll.
XX Photorhabdus luminescens W-14; nematode; symbiotic;
XX Heterorhabdus; tcd; tcb; tcc; tcd; insecticidal activity; toxin;
XX Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
XX Homoptera; Southern; Western corn rootworm; Colorado potato beetle;
XX mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;
XX cabbage looper; codling moth; corn earworm; European corn borer;
XX Tobacco hornworm; budworm; ds.
XX Photorhabdus luminescens.
OS
XX Key Location/Qualifiers
XX CDS 1..5547
XX /*tag= a
XX /note= "no stop codon given"
XX
XX W09808932-A1.
XX
XX 05-MAR-1998.
XX
XX 05-MAY-1997; 97MO-US07657.
XX
XX 06-NOV-1996; 96MO-US18003.
XX 28-AUG-1996; 96DS-0705484.
XX 06-NOV-1996; 96DS-0743699.
XX
XX (DMC) DOWELANCO.
XX (WISC) WISCONSIN ALUMNI RES FOUND.
XX
XX Blackburn MB, Bowen DJ, Cliche TA, Ensign JC, Fattig R;
XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GL;
XX Pettell J, Roberts JL, Rochelleau TA, Schoonover S;
XX Strickland JA, Sukhplinda K;
XX
XX WPI; 1998-179427/16.
XX P-PSDB; W56573.
XX
XX Isolated toxins from Photorhabdus luminescens strains - useful for
XX control of insect pests
XX
XX Claim 38; Pages 231-237; 321pp; English.
XX
XX The present sequence encodes a protein named TcdAll of the bacterium
XX Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
XX nematodes of the Heterorhabdus genus. The bacterium has at least 4
XX distinct genomic regions, tcd, tcb, tcc, and tcd. Peptide products are
XX produced from these regions that are associated with insecticidal
XX activity. The native toxins are secreted proteins. The proteins are

```

```

CC toxic to insects upon exposure and especially when ingested. The
CC nucleic acid sequence can be used to produce transgenic plants,
CC baculoviruses or microbial hosts for toxin production. They can be used
CC to control insects pests from the lepidoptera, Coleoptera, Hymenoptera,
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer or tobacco hornworm
CC or budworm.
XX
XX Sequence 5547 BP; 1711 A; 1223 C; 1112 G; 1501 T; 0 other;
SQ
alignment_scores:
Quality: 152.00 Length: 809
Ratio: 0.420 Gaps: 37
Percent Similarity: 44.747 Percent Identity: 18.294
alignment_block:
US-09-494-297-2 x V29930 ..
Align seg 1/1 to: V29930 from: 1 to: 5547
157 AlavAlMetYrAsnGlyHsProGlnAsnAlaAsnGlyIleMetGlu 173
173 YleuGluProLeuAsnAlaIleArgValThrGlnGlu..AlavAlTpt 189
2433 GCtGATATATTCATCATGAAAGACACCGACCTATGCCAGTGGG 2482
189 YrYrSerAsp.....AsnAlaProIleSerAsn 198
2483 AAAACGCGCAGCGGTATTAACCGCGCGGTGAATTCACAAAGCGTAA 2532
199 ProAspGluSerPheLysArgGluSerGluSerAsnLeuValSerThr 214
2533 ACATTAACGCTTTCTGCGATGATTCGCGATGCGGATTAAGCACTA 2582
215 .....SerGlnLeuSerLeuMetArgGlnAlaLeuLysGln..... 226
2583 CTATATCCGTCAAGTCCGCAAGCAGCGCGGTATTAAGCGCGTATG 2632
227 .....LeuIleAspProAsnLeuAla..... 233
2633 ACTTGATCAATACTTACTGATGATATATCAGGTTTCTGCGCAATAAA 2682
234 ...ThrLysMetProLysGlnValProAspAspPheGlnLeu..... 246
2683 ACCACCGCGATCGCGCAAGCCATTTGCC...AGTATTCACTGATCGTCA 2729
247 .....SerIlePheGluSerGluAspLysGly..... 255
2730 CCGGGCATTTGGAAATATGGAAGAAATGCAATTCGGGGTTATCAGCC 2779
256 .....AspLysTyrAsnLysGlyTyrGlnAsnLeu 265
2780 GCCAATTTCTTATGACTGCGCAAAATACATAAACCCTCAGACGTTGG 2829
266 LeuSerGlyGlyLeuValProThrLysProProThrProGlyAspProp 282
2830 CGGGGTGTTTCAATTAGTTTACTACCGGAACATATATGATCCGAC 2879
282 oMetProProAsnGlnProGlnThrThrSerValLeuIleArgLysTyrA 299
2880 CATCGTATGCGACAAACCAAAATGATGAGCATTTACTGCAATCGCTCA 2929
299 la..... 299
2930 GCCAAAGCAATTAAGCCGATACGTCGAAGATGCCCTTATGCTTAT 2979
300 IleGlyAspYrSerLysLeuLeuGluGlyAlaThrLeuGlnLeuThrG 316

```



2980 CTGACATCGTTTGAAACAAGTGGCTAATCTTAAAGTTATTACGCATATCA 3029  
 316 YASPAStValAsnSerPheGlnAlaArgValPhe.....Sers 329  
 3030 CGATAAATTAATTAACATCAAGGGCTAGCTATTATTATCGACTCAGTG 3079  
 329 eRAsnAspIleGlyLys.....ArgIleGluLeu 338  
 3080 AACTGATGCCGGGTAAATATTATTGGCGCAGTGCATCACAGTAATATTC 3129  
 339 SerAspIleThrThrLeuThr.....GluLeuAs 349  
 3130 AAGCAGCGTAATTCGGCGCAATGCGTGGAGTGAATGGCATAAATTTGA 3179  
 349 nSerProAlaGly.....TyrSerIleAlaGluProIleThrPheLysV 364  
 3180 TTGTCCATTAACCCCTATTAACACACTATCCGTCAGTATATATAAA. 3228  
 364 alGluAlaGlyLysValThrIle...IleAspGlyLysGlnIleGlu 379  
 3229 .....TCCCGCTGATCTGCTGCTGTGGTGGAAACAAGAGATCC 3270  
 380 AsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAsnAs 396  
 3271 AAACAGACAGGAATAGTAAAGATGGCTATCAACTGAACGATTAATCG 3320  
 396 pPheGlu..... 398  
 3321 TTATGAACATAAATTGGCGCATATCCCTATGATGACACTTGAATACGC 3370  
 399 .....GluPheSerValLeuThrThrGlnAsn 407  
 3371 CAATCACCCTTGATGATCAATAAATAATCCGAGCTAAACCTGAAAAA 3420  
 408 .....TyrAlaLysPheTyrTyrAlaLysAsn..... 416  
 3421 AATAGAGCGCCCGGAGCTATATTGTCCGGTTATCAAGGTGAAGTACGTT 3470  
 417 .....LysAsnG 419  
 3471 GCTGGTATGTTTATAACCAACAAGACACACTAGATAGTTATTAATAAGC 3520  
 419 LysSerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProPro 435  
 3521 CTTCATGCAAGGACTATATATCTTT...GCTGATGCGATCC..... 3561  
 436 AspSerGluAspGlyGlyLysThrMetThrPro..... 446  
 3562 .....AAGATATGACCCCAAGACAGACATATGTTTA 3593  
 446 ..... 446  
 3594 TCGGATTAATAGCTATCAACAATTGTATACCAATATATGTCAGAAAGTGA 3643  
 446 ..... 446  
 3644 ATAAACCGCTATGACAGAGATTATAGATCTCTCCGCGTAAGTACCGCT 3693  
 447 ...AspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAs 462  
 3694 AAGAGCTATGTTGGGAGATTATTAACCTGACATGCTATTAACGGAGA 3743  
 462 PleuPheLysTyrThrValLysProArgAspThrAspProAspThrPheL 479  
 3744 TATTCCAACTTCAATTAACAAGCCGATCAAGATTTAAAAATCTATA 3793  
 479 euLysHisIleLysValIleGluLysGlyTyr.....Arg 491  
 3794 TCTGACCAAAATTAAGATTAATCATATATGATGAAAGACAGACAGCGC 3843  
 492 GluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGlnIleuArg 508  
 3844 AATCAATGCAATCTGATGAATTAATATGCGCAAACTAGCTAATTAATTAT 3893

508 gAlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuA 525  
 3894 TGTATTATCTAGCTTGGGGTCAATTCAAATTAACCTGTCATTAATAGCTCA 3943  
 525 sPLysAspLysLeuLysAspTyrHisGlyPheGlyAspMetAsnAspSer 541  
 3944 TGTTTTACCCGCTCTATCAATATAGCGA.....AACACCGAT 3981  
 542 ThrLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAspSerAsnPr 558  
 3982 GGACTCAATCAAGGAGACTACTA.....TTCACCGCTGACACACTTA 4025  
 558 oProGlnLeuThrAspLeuAspPheIleProAsnAsnLys..... 573  
 4026 TCCA.....TCTAAAGTAGAAGCTTGATTCCTGGACCAAAAGCTTCTC 4069  
 574 .....TyrGlnSerLeuIleGlyThrGlnTrp..... 582  
 4070 TAACCAACCAAAATGCCCGCATTTGTGATGATTAATGCTACAGACTCTCTG 4119  
 583 ...HisProGluAspLeuValAspIleIleArgMetGluAspLysGly 598  
 4120 AATAAACCGATGATCTTAACCAATATATCTTATGACTGACAGTAAAG 4169  
 598 uValIle.....ProValThrHisAsnLeuThrLeu... 608  
 4170 GACGTACTAGTATGCTCAGGCCAGTAGAGATTAATATCTCAATTTCTC 4219  
 609 .....ArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPhe 623  
 4220 CAGCAAAAGTTCAGTATATAGTCAAAAGCGGCTGCCAAGACCAAACTTTT 4269  
 624 HisPheGlu..... 626  
 4270 ACCGAGATTAAGATGCTCCATTCAGCCATCACCTACTTGTATGAAT 4319  
 627 .....IleGluLeuLysAsnAsnLysGlnGluLeu 637  
 4320 GAATTAATCAATTAATGCGCTGAAATAGACGGTCTGCTGATTAATTA 4369  
 637 euSerGlnThrValLysThrAspLysThrAsnLeuGluPhe...LysAsp 652  
 4370 TTAAACACTGACCGCAGTATGATGTTACTTTTACCGCATTTGGCGAGCAT 4419  
 653 GlyLys.....AlaThrIleAsnLeu 659  
 4420 GCGCGCAAAAGTGGTTATGAAGTTTCAGTATTCCTGTACCCCTCAAGST 4469  
 659 uLysHisGlyLysLeuThrLeuGlnGlyLeuProGluGlyTyrSerT 676  
 4470 AAGTACCGATATGCGCTGACCCCTGACCAATTAAGAAATGCTGGCAAT 4519  
 676 YrLeuValLysGluThrAspSerGluGlyTyrLysValLysValAsnSer 692  
 4520 ATATG.....CAATGGCAATCTTATCGTACCCGCTGAATACT 4557  
 693 .....GlnGluValAlaAsnAlaThrValSerLysThrGlyI 705  
 4558 CTATTGCCCCCGAGTGTGTCACGCCGACAC.....ACCGAAT 4598  
 705 eThrSerAspLeuThrLeuAlaPheGluAsnAsnLysGluProValValP 722  
 4599 CGATCAATTTCTGATGAGAACTCAGAAATATTCAGGACCG..... 4641  
 722 roThnGlyValAspGlnLysIleAsnGlyTyrLeuAlaLeuIleValIle 738  
 4642 .....CAGTTAGCAAAAGTTTCTATGCTACTAGCTCTGATA 4677  
 739 AlaGlyIleSerLeuGlyIleTyrGly 747  
 4678 CTTCCCTATTAACCTATCAACATCATGCT 4704



```

3493 .....TCCGCGCTGATCTGCTGCTGTGGAGAACAAAGAGATCCGC 3534
380 AsnProAsnLysGluIleValGluProTyrSerValGluAlaTyrAsnAs 396
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3535 AATCAGACAGAAATAGTAAGATGGCTATCAAACTGAAGAGATTATCG 3584
396 pPheGlu..... 398
3585 TTATGAACATAAATTTGGCGCATATCCGCTATGATGGCATTGGAATACGC 3634
399 .....GluPheSerValIleuThrThrGlnAsn 407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3635 CAATCACCCTTGATGTCATATAAAATATCCGACTTAACCTGGAATAAA 3684
408 .....TyrAlaLysPheTyrTyrAlaLysAsn.... 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3685 AATAGAGCGCCGCGACTTATTTGGCGGTATCAAGGTGAAGATACGTT 3734
417 .....LysAsnG 419
3735 GCTGCTGATGTTTATACCAACAAGACACTAGATATTAATAAAACG 3784
419 LysSerGlnValValTyrCysPheAsnAlaAspLeuLysSerProPro 435
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3785 CTTCATGTCAGAGACTATATATCTTT...GCTGATATGGCATCC..... 3825
436 AspSerGluAspGlyGlyLysThrMetThrPro..... 446
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3826 .....AAGATATGACCCCGAGAACAGCAATGTTTA 3857
446 ..... 446
3858 TCGGATATAGCTATCAACAATTGAGACCAATATGTCAGAAAGTGA 3907
446 ..... 446
3908 ATACCCGTATGACAGAGATTATGATCTCTCCGCTGATAGACCGT 3957
447 ...AspPheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAs 462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3958 AAGACTATGCTGGGAGATATTACTCAGCATGCTATTAACGAGGA 4007
462 PheUphelThrValLysProArgAspThrAspProAspThrPheL 479
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4008 TATTCACACTATCAATTACAAAGCCGCAATCAAGTATTAATAATCTATA 4057
479 euLysHisIleLysValIleGluLysGlyTyr.....Arg 491
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4058 TCTCACCAAAATTAAGATTAATTCATTAATGATATGAAGACAGAGACGC 4107
492 GluLysGlyGlnAlaIleGluTyrSerGlyLeuThrGluThrGlnLeuAr 508
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4108 AATCATGTCATGATGATTAATATGCGCAACTAGCTGATTAATTTAT 4157
508 GAlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuA 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4158 TGTATTACTGCTGGGGCTCAATCCAAATACTCGTCAATAATAGTCGA 4207
525 sPlysAspLysLeuLysAspTyrHisGlyPheGlyAspMetLysAspSer 541
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4208 TGTTTTACCCGCTATCAATATAGCGGA.....AACACCGGT 4245
542 ThrLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAspSerAspR 558
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4246 GGACTCATCAAGGAGACTACTA....TTCACCCCTGACACCACTTA 4289
558 oPrgGlnLeuThrAspLysAspPhePheIleProAsnAsnLys.... 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4290 TCCA.....TCTAAAGTAGAAGCTTGATTCCTGAGACAAACGTTCTC 4333
574 .....TyrGlnSerLeuIleGlyThrGlnTP..... 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

4334 TAACCAACCAAAATGCGCGCATGTGGTATGATTTATGCTACAGACTCTCTG 4383
583 ...HisProGluAspLeuValAspIleIleArgMetGluAspLysGly 598
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4384 AATTAACCGGATGATCTTAACCAATATGATCTTTATGACTGACAGTAAGG 4433
598 uValIle.....ProValThrHisAsnLeuThrLeuArgL 610
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4434 GACTGCTACTGATGCTCAGGCGCAGTAGAGATTAATATCTGCAATTTCTC 4483
610 yThr.....ValThrGlyLeuAlaGlyAspArgThr... 620
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4484 CAGCAAAAGTTCAGATATATAGTCAAAAGCGGTGGCAGACAGCAAACTTT 4533
621 .....LysAspPheHisPheGlu..... 626
4534 ACCGCGATTAAGAATGTCCTCATTCAGCCATCACCTAGCTTGTATGAAT 4583
627 .....IleGluLeuLysAsnAsnLysGlnGluLeuL 637
4584 GAATTAATCAATTAATGCGCTGAAATAGACGGTTCTGCTGAATTTTA 4633
637 euSerGlnThrValLysThrAspLysThrAsnLeuGluPhe...LysAsp 652
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4634 TTAACTACTCAGCCAGTATGATGTTACTTTACCGCATTTTGGCGAGAT 4683
653 GlyLys.....AlaThrIleAsnLe 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4684 GGCGCAAAAGTGGTTATGAAGTTTCAGTATCTCTGTTACCCCTCAAGT 4733
659 uLysHisGlyLysSerLeuThrLeuGlnGlyLeuProGlnGlyTyrSer 676
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4734 AAGTACGATATGCTGACCTGCACCATTAATGAATAATGTCGCCAAT 4783
676 yLeuValLysGluThrAspSerGlnGlyTyrLysValLysValAsnSer 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4784 ATATG.....CATGGCAATCTATCGTACCCGCTGATATCT 4821
693 .....GlnGluValAlaLysAlaThrValSerLysThrGlyI 705
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4822 CTATTTGCCCGCCAGTTGTTGACAGCCGCCAC.....ACCGAAT 4862
705 eThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValAlP 722
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4863 CGATACAACTGCTGATATGAGAACACAGAAATATCAGAACCG..... 4905
722 roThrGlyValAspGlnLysIleAsnGlyTyrLeuAlaLeuIleValIle 738
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4906 .....CAGTTAGCAAAAGGTTTCTATGCTACGTTCTGATA 4941
739 AlaGlyIleSerLeuGlyIleTyrPgly 747
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4942 CTTCCCTTAATCACTATCAATCATGCT 4968
seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT.V29928
seq_documentation_block:
ID: V29928 standard; DNA; 7551 BP.
XX
AC V29928;
XX
DT 07-AUG-1998 (first entry)
XX
XX
DE tcd gene from the tcd genomic region encoding protein TcdA.
XX
KW Photorhabdus luminescens W-14; nematode; symbiotic;
KW heterorhabditis; tcd; tcd; tcd; insecticidal activity; toxin;
KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;
KW Homoptera; Southern corn rootworm; Colorado potato beetle;
KW mealworm; boll weevil; turf grub; beetle armworm; black cutworm;
KW cabbage looper; codling moth; corn earworm; European corn borer;
KW tobacco hornworm; budworm; ds.
XX

```



```

3685 AATAGAGCCGCCGACTATGTGCGGTTATCAAGTGAAGTACGTT 3734
      |||::: ||| :::::
417 ..... LysAsnG 419
3735 GCTGGTATGTTTATACCAACAAGACACATGATATGTTATAAAGC 3784
      |||::: |||:::
419 LysSerSerGlnValValTyrCysPheAsnAlaAspLeuLysSerPro 435
      |||::: |||::: |||::: |||:::
3785 CTTCATATGCAAGAGACTATATATCTT...GCTGATATGGCATCC... 3825
436 AspSerGlnAspGlyGlyLysThrMetThrPro... 446
      ||| ||| ||| |||
3826 ..... AAGATATGACCCCAAGACAGACATGTTTA 3857
446 ..... 446
3858 TCGGGATATATAGTATCAACAATTGTATACCAATATGTAGAGAGTGA 3907
446 ..... 446
3908 ATAAACCGCTATGAGAGATTATGAGATTCTCTCGGTAAAGTACCGT 3957
447 ...AspPheThrGlnGlyGluValLysTyrThrHisIleAlaGlyArgAs 462
      |||::: |||::: |||::: |||:::
3958 AAAGACTATGGTTGGGAGATTATTACCTCAGCATGGTATATACGAGA 4007
462 pLeuPheLysTyrThrValLysProArgAspThrAspProAspThrPheL 479
      |||::: |||::: |||::: |||:::
4008 TATTCACACTATCAATACAAAGCCGATCAAGATTTAAATATCTATA 4057
479 euLysHisIleLysLysValIleGluLysGlyTyr...Arg 491
      |||::: |||::: |||::: |||:::
4058 TCTCACCACAAATTAAAGATATTCATATGATATGAAGACAGAGAGCGC 4107
492 GlnLysGlnAlaIleGluTyrSerGlyLeuThrGlnThrGlnLeuAr 508
      |||::: |||::: |||::: |||:::
4108 AATCAATGCAATCTGATGATTAATATATGCAACTAGTGTATTAATTTAT 4157
508 gAlaAlaThrGlnLeuAlaIleTyrTyrPheThrAspSerAlaGluLeuA 525
      |||::: |||::: |||::: |||:::
4158 TGTATTACTAGTCTGGGGGTCATCAAAATACTGTCGCAAAATAGTCA 4207
525 sPlyAspLysLeuLysAspTyrHisGlyPheGlyAspMetAsnAspSer 541
      |||::: |||::: |||::: |||:::
4208 TGTTTACCCGCTATCAATATAGCGA...AACACACAGT 4245
542 ThrLeuAlaValAlaLysIleLeuValGluTyrAlaGlnAspSerAsnPr 558
      |||::: |||::: |||::: |||:::
4246 GGACTCAATCAAGGAGACTACTA...TTCACCGCTGACACCACTTA 4289
558 oProGlnLeuThrAspLeuAspPheIleProAsnAsnLys... 573
      |||::: |||::: |||::: |||:::
4290 TCCA...TCTAAGTAGAAGCTTGATTCCTGAGCAAAAGCTTCTC 4333
574 .....TyrGlnSerLeuIleGlyThrGlnTP... 582
      |||::: |||::: |||::: |||:::
4334 TAAACCAACCAAAATGCCGCTATGATGTATGCTATACAGACTCTCTG 4383
583 ...HisProGlnAspLeuValAspIleIleArgMetGluAspLysGln 598
      |||::: |||::: |||::: |||:::
4384 AATTAACCGGATGATCTTAACGATATATCTTTATGACTGACAGCTAAGG 4433
598 uValIle...ProValThrHisAsnLeuThrLeu... 608
      |||::: |||::: |||::: |||:::
4434 GACTGCTACTGATGTCACGGCCAGTAGAGATTAATATCTGCAATTTCTC 4483
609 ...ArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPhe 623
      |||::: |||::: |||::: |||:::
4484 CAGCAAAAGTTCAAGATATAGTCAAAAGCGGTGCGCAAGAGCAAACTTTT 4533
624 HisPheGlu... 626
      |||::: |||::: |||::: |||:::

```

```

4534 ACCGAGATTAAGATGTCCTCATTCAGCCATCACCCTAGCTTGTGAAT 4583
627 .....IleGluLeuLysAsnAsnLysGlnGluLeu 637
      |||::: |||::: |||::: |||:::
4584 GAATTATCAATTTAATGCGCTTGAATATGAGCGTTCTGCTGTAATTTTA 4633
637 euSerGlnThrValLysThrAspLysThrAsnLeuGluPhe...LysAsp 652
      |||::: |||::: |||::: |||:::
4634 TTAACAACTCCAGCGATGATGTTACTTTTACCGCATTTGGCGAGGAT 4683
653 GlyLys...AlaThrIleAsnLe 659
      |||::: |||::: |||::: |||:::
4684 GGCCGCAAACTGGGTATGAAGTTTCAGTATCTCTTACCTTACCTCAAGT 4733
659 uLysHisGlyGlnSerLeuThrLeuGlnGlyLeuProGlnGlyTyrSerT 676
      |||::: |||::: |||::: |||:::
4734 AAGTACGATATATGCGCTGACCTGACCATATGAATAAATGTGCGCAT 4783
676 YrLeuValLysGluThrAspSerGluGlyTyrLysValLysValAsnSer 692
      |||::: |||::: |||::: |||:::
4784 ATATG...CAATGGCAATCTTATCGTACCGGCTGAAATACT 4821
693 .....GlnGluValAlaAsnAlaThrValSerLysThrGlyTl 705
      ||| ||| ||| ||| ||| ||| ||| |||
4822 CTATTTGCCCGCCAGTGTGTCACGCCGCCAC...ACCGAAT 4862
705 eThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValValP 722
      |||::: |||::: |||::: |||:::
4863 CGATACAAATTCGTAGTATGCAAACTCAGATATATCAGAACCG... 4905
722 roThrGlyValAspGlnLysIleAsnGlyTyrLeuAlaIleValIle 738
      |||::: |||::: |||::: |||:::
4906 .....CACTTACGCAAAAGGTTTCTATGCTACGCTGCTGATA 4941
739 AlaGlyIleSerLeuGlyIleTyrGly 747
      |||::: |||::: |||::: |||:::
4942 CTTCCCTTATACCTTATCACTCATGCT 4968
seq_name: /cgnl_8/gc9data/geneseq/geneseqn/NA1999.DAT:X77451
seq_documentation_block:
ID X77451 standard; DNA; 3531 bp.
XX
AC X77451;
XX
DT 05-AUG-1999 (First entry)
XX
DE S. pyogenes SFFBP-12 DNA.
XX
KW SFFBP-12; fibronectin; fibrinogen; group A Streptococci; infection;
KW fibrinogen and fibronectin binding protein; bacterial adhesion; vaccine;
KW diagnosis; treatment; prevention; streptococcal infection; antigen;
KW immune system; etiologic agent; suppurative infection; pharyngitis;
KW impetigo; necrotizing fasciitis; systemic disease; scarlet fever;
KW toxic-shock syndrome; sequelae; rheumatic fever; glomerulonephritis; ss.
XX
OS Streptococcus pyogenes.
XX
PN US910441-A.
XX
PD 08-JUN-1999.
XX
PF 16-SEP-1996; 9605-0714402.
XX
PR 16-SEP-1996; 9605-0714402.
XX
PA (UVRQ ) UNIV ROCKEFELLER.
XX
PI Fischetti VA, Rocha C;
XX
DR WPI: 1999-356822/30.
DR P-PDB; Y08603.
XX

```



```

324 AlaArgValPheSerSerAspIleGlyGlu.....ArgIleGluLe 338
1819 AAATAGACAGCCAGTTCACACACAGAGAGAGTTCACCTTAAGAGACT 1868
338 userAspIleThrThrLeuThrLeuAsnSerProAlaGlyTyrS 355
1869 GACCTCGGCGACATATGACCTGTATGAAACAAAGCCCAAGAGTTATC 1918
355 er..... 355
1919 AGCAGGTGACAGAGAAATGGCGACCGTTACTGTGATACCAAACT 1968
356 IleAlaGluProIleThrPhe.....LysValGI 365
1969 CCTGAGAAATGGTCACCTTGGGGAAGCCCATTCGTCTGTAAGAATGA 2018
365 uAlaGlyLysValTyrThrIleIleAspGlyLysGln...IleGluAsn 381
2019 AGCTAACAAAGAGTCACGATTTGCACACCAAAAGAACCTTACGTTT 2068
381 roAsnLysGluIleValGluProTyrSerValGluAlaTyrAsnAspPhe 397
2069 CAGGAGAGAAATTTGGGAG.....AATGACAGA 2097
398 GluGluPheSerValLeuThrThrGlnAsnTyrAlaLysPheTyrTyrAl 414
2098 CCAAT.....CAACGCCACCAAGATTCAGTCA 2129
414 alyAsnLysAsnGlySerSerGlnValValTyrCysPheAsnAlaAspL 431
2130 ACTGTGCAAAATGTCA..... 2148
431 euLysSerProProAspSerGluAspGlyLysThrMetThrProAsp 447
2149 ..AAGATCCCTAACCCAGATTCAGAA.....GTAAAGAGAGAT 2184
448 PheThrThrGlyGluValLysTyrThrHisIleAlaGlyArgAspLeuP 464
2185 .....AAGCATGGCTTAT...CACTTC.....AAGACTTCC 2216
464 eLysTyrThrValLysProArgAspThrAspProAspThrPheLeuLysH 481
2217 TAAGTACCATCCCAAGATCAGAG..... 2241
481 IsIleLysLysValIleGluLysGlyTyrArgGluLysGlyAlaIle 497
2242 ..TATAGTACTCAGTTGAGAGAAATGTTCCAGACGCGTCAAGGTG 2289
498 GluTyrSerGlyLeuThrGluThrGlnLeuArgAlaIleThrGlnLeuAl 514
2290 TCGATATTTAGGAATGATATTTAAACCCAGA..GAACAGAAATTTGT 2336
514 aileTyrThrPheThrAspSerAlaGluLeuAspLysAspLysLeuLys 531
2337 GTTTCACACAGATACTTAACCTTGAATTTGGAAATCTGAATAATAAAG 2386
531 sPTyrHisGlyPheGlyAspMetAsnAspSerThrLeuAlaValAlaLys 547
2387 GTCATCTGGGTCAAAATCATTTGATGAAGACAGCTAACGCTTTCAA 2436
548 IleLeuValIleTyrAlaGlnAspSer.....AsnProProGlnLeuTh 562
2437 GGTAAAGAAATTTGGAAAATGATATACGACAGAAATCTCCCAAGCAT 2486
562 rAspLeuAspPhePhe.....IleProAsnAsnLysT 574
2487 TCAAGTGCACCTTATCTGATGAGATGCGCTGTGAGAGGTCAACCAAT 2536
574 yrgInSerLeuIleGlyThrGlnTyrHisProGlu..... 585
2537 TTATTTCTGGCTCAGGTAAATGATGTCATTTGAGTTAAAAACTTGAAG 2586
586 .....AspLeuVal..... 588

```

```

2587 AAGTATATGAGACAGTAAATGACATCATTTACTACGTTAAAGAAGTAC 2636
589 .....AspIleIleArgM 593
2637 TGTTCACACAGATTATGATGTGACTTACTACGTAATGATATTTAT... 2682
593 etGluAspLysLysGluValIle.....ProVal 602
2683 ..AATACCAAAAGTGGTATTTACACAAAGAGCCGAAACCTAGAGATT 2730
603 ThrHisAsnLeuThrLeuArgLysThrValThrGlyLeuAlaGlyAspAr 619
2731 GAAGAAACGCTTCGCTAGAAATCAGAGTCTTCAGC.....GGTACAC 2774
619 gThrLysAspPheHisPheGluIleGlu.....LeuLysAsnA 632
2775 TACTTCGAAGACTCAGCCAGTTGATACCTTATACAGTTTATCAAGTG 2824
632 snLysGlnGluLeuLeuSerGlnThrValLysThrAspLys...ThrAsn 647
2825 AGCAAGTCACTCCGCTGATATGACAAATTTGAAGAGATGCTACCAT 2874
648 LeuGluPheLys.....AspGlyLys.....AlaTh 656
2875 ATTATATTCATAAAGCTGATATTGACGCGCAAGAGTTAGCTGGTGCAAC 2924
656 rIleAsnLeuLysHisGlyGluSerLeuThrLeuGlnGly..... 669
2925 TATGAGATTGGCTGATTCATCTGTAATAATATATGATCATGATTCAG 2974
670 .....LeuProGluGlyTyrSerTyrLeu 677
2975 ATGCAAGTGAAGATTTCACCTGATGACGCAAGAAATATACATTT... 3021
678 ValLysGluThrAspSerGluGlyTyrIleVal..... 688
3022 GTCCAAACCCGACGACCCAGACGGTTATGATAGTACACTGCTATTACCT 3071
689 ..LysValAsnSerGlnGluValAlaAsnAlaThrValSerLysThrGlyI 705
3072 TACAGTTATATGAGCA.....GGTCAGGTACTGTAATGCAAAAGCA 3115
705 IeThrSerAspGluThrLeuAlaPheGluAsnAsnLysGluProValVal 721
3116 CTAAAGGTGACACTCATATGTCATGTTGATGCTTACAAAGCCAACTAAG 3165
722 ProThrGly.....ValAspGlnLysIle 729
3166 GGTTCAGGTCAAGTTATGATTTGAAGAAAGGCTT 3201
seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT:A97541
seq_documentation_block:
ID A97541 standard; cDNA, 2700 BP.
XX
AC A97541;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human Hsp72 (heat shock protein 72) cDNA.
XX
KW Human Hsp72; heat shock protein 72; chromosome 6p21.3; Hsp72 inhibitor;
KW expression modulator; JNK phosphatase inhibitor; antiproliferative;
KW drug screening; cancer; leukemia; lymphoma; solid tumour; sarcoma;
KW carcinoma; breast cancer; prostate cancer; premalignant condition; ss.
XX
OS Homo sapiens.
XX
PN MO200054814-A1.
XX
PD 21-SEP-2000.
XX

```

```

PF 17-MAR-2000; 2000WO-US07350.
XX
XX 18-MAR-1999; 99US-0125046.
XX
XX (PHYL-) PHYLGENY INC.
XX
XX Volloch VZ, Sherman M;
PI
DR WPI: 2000-647056/62.
DR P-PSDB: B23252.
XX
XX Identifying compounds that inhibit proliferation of cells and capable
PT of modulating the expression of heat shock protein 72 gene and/or
PT activity of Hsp72 useful for treating cancers such as leukemia,
PT lymphoma
XX
XX Examples: Fig 16a: 77pp: English.
XX
XX The invention relates to a novel method of identifying compounds that
CC inhibit proliferation of cells comprising contacting a test compound with
CC a cell which overexpresses Hsp72 (heat shock protein 72), and determining
CC if the test compound inhibits activity or expression of Hsp72.
CC Optionally, Hsp72 is contacted with the test compound under optimum
CC conditions to allow the two components to interact and bind, forming a
CC complex which is detected. The invention also relates to a method of
CC identifying compounds that inhibit Hsp72-mediated JNK phosphatase
CC activation, comprising contacting a test compound with a cell which
CC expresses Hsp72, exposing the cell to a heat induced stress and
CC determining if the compound inhibits JNK phosphatase activity. The
CC invention additionally encompasses compositions comprising an inhibitor
CC of Hsp72 or JNK phosphatase activity. The compounds identified as
CC inhibitors of Hsp72 or JNK phosphatase activity are useful for
CC inhibiting the proliferation of cells. Modulation of the activity of the
CC JNK phosphatase or Hsp72 is used to treat a proliferative disorder such
CC as cancers (e.g., leukemia, lymphoma, solid tumours such as sarcomas and
CC carcinomas, breast cancer, prostate cancer). The compounds that inhibit
CC Hsp72 activity can also be administered to treat premalignant conditions
CC and to prevent progression to a neoplastic or malignant state. The
CC compounds that inhibit Hsp72 function are administered to a patient
CC having a disease or disorder mediated by an increase of Hsp72 expression
CC or activity relative to normal levels. The present sequence represents
CC cDNA encoding human Hsp72 used in the exemplifications of the invention.
XX
XX Sequence 2700 BP; 601 A; 780 C; 834 G; 485 T; 0 other:

```

```

alignment_scores:
    Quality: 149.00      Length: 655
    Ratio: 0.463        Gaps: 34
Percent Similarity: 49.160 Percent Identity: 22.137

```

alignment\_block:

US-09-494-297-2 x A97541 ..

Align seg 1/1 to: A97541 from: 1 to: 2700

```

141 AlMeSerProArGileTHngLYASpGLuLeuAsnGLnLysleuArgAl 157
    ||| |||
189 GCCTCATCGAGCTCGGTGATTGGCTCAAGGGAAGGCGGGCTCCGT 238
    |||
157 aValMetTyAsn.....GlyHisProGlnAsnAlaAsnGlyTleMetG 172
    : : : : : |||
239 GAGGACTTATAAAGCCAGGCGGCAAGCGGTACGGGCTAGCGCTG 288
    |||
172 lueGluLeuGluProLeuAsnAlaIleArgValThrGlnGluAlaValTyr 188
    |||
289 AGGAGCTG.....CTGCACAGATCCACATCACTT 317
    |||
189 TyrTyrSerAspAsn...AlaProIleSerAsnProspGluSerPhe 204
    : : : : : |||
318 TTCAGAGTACTCCGTTGCCAAGGCTTCCAGAGCGAACCCTGTGCG 367
    |||
204 sarGluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeuMetA 221

```

```

: : : : : |||
368 GCTGCAGGACACCGCGCTGAGTTTCCGCC.....GTCC 402
    |||
221 rglAlaIleuLysGlnLeuIleAspProAsnLeuAlaThrLysMetPro 237
    |||
403 GGAGGACCGAGCTCTTCTCGGATCAGTGTTCGTTCCAGCCCCCA 452
    |||
238 LysGlnValProAspAspPheGlnLeuSerIlePheGluSerGluAsp 254
    : : : : : |||
453 ATCCACAGCGGAGCCGACAGAGACGAGACCGGATGCCAAGCCG 502
    |||
254 sGlyAspLysTyraAsnLysGlyTyrGlnAsnLeuLeuSerGlyGly 271
    : : : : : |||
503 CGCGGATCGGCAATGACCTGGGACACCACTACTCTCTCGGGGGGT...G 549
    |||
271 alProThrLysProProThrProGlyAspProPheMetProProAsnGln 287
    |||
550 TTCACACGCG.....CAAGGTGAGATCATCGCCACAGCACGAGGCAA 593
    |||
288 ProGln.ThrThrSerValLeuIleArgLysTyrAlaIleGlyAspTyr 304
    |||
594 CCGGACACCCCGCAGCTACGTG.....GCCTTACGGGACACCG 631
    |||
304 eryLysLeuLeu...GluGlyAlaThrLeuGlnLeuThrGlyAspAsnVal 319
    : : : : : |||
632 AGCGGCTATCGGGGATCGCGCAAGAACAGGTCGCTGAACCCCGCAG 681
    |||
320 AsnSer...PheGlnAlaArgValPheSerSerAsnAspIleGlyGlu 335
    : : : : : |||
682 AACACCGTGTTCAGCGGAGGCGGCTATTGGCGCAAGTTCGGCGGCC 721
    |||
335 gIleGluLeuSerAspGlyThrTyr.....ThrLeuThrGluL 348
    |||
732 GGTGTGCTAGTCGACATGAGACACTGCTTCCAGGTATCAACAGCAG 761
    |||
348 euAsnSerProValaglyTyrSerIleAlaGluProIleThrPheLysVal 364
    : : : : : |||
782 GAGCAAGGCC.....AAGTGCAGGAGACTCAACAGGG 816
    |||
365 GluAlaGlyLysValTyrThrIleIleAspGlyLysGlnIleGluAsn 381
    : : : : : |||
817 GAGCCAAAGCATTCATCCCGGAGAGATCTGTCATGTGTGTGCTGACCA 866
    |||
381 oAsnLysGluIleValGluProTyr..... 389
    |||
867 GATGACGAGATCGCCGAGCGGTACTGGGCTACCGGTACCAACAGCGG 916
    |||
390 .....SerValGluAlaTyr...AsnAspPheGluGlu..... 399
    : : : : : |||
917 TGATCACCCTGTCGGCCTACTTCAACGACTCGCAGCGCAGCCACCAAG 966
    |||
400 .....PheSerValLeuThrThrGlu 406
    : : : : : |||
967 GATCGGGGTGTATCGGGGCTCAACGTCGTGGGATCAACAGCAGCC 1016
    |||
406 nasuTyraIalysPheTyraIalysAsnLysAsnGlySerSerGln. 422
    : : : : : |||
1017 CACGGCCGCCCATCGCTACGCGCTGACAGACAGCGGCAAGGGGAGC 1066
    |||
423 ...ValValTyrCysPheAsnAlaAspLeuLysSerProProAspSerGlu 438
    : : : : : |||
1067 GCAACGTCATCACTTT.....GACCTG..... 1089
    |||
439 AspGlyGlyLysThr.....MetThrProAspPheThr 450
    : : : : : |||
1090 ...GGCGGGGCACTTCGAGCTGTCATCTCACTGACATGACAGGCGAT 1136
    |||
450 rGlyGluValLysTyr.....ThrHisIleAlaGlyArgAspL 463
    : : : : : |||
1137 CTTCAGAGTGAAGCCAGCGCGGAGACACCACTGGGTGGGAGAGACT 1186
    |||
463 eupheLysTyrThrValLysProArgAspThrAspProAspThrPheLeu 479
    : : : : : |||

```



```

1187 TTCACACAGGCTGTC.....AACCACTGCTG 1215
480 LysHisIleLysValIleGluLysGlyTyrArgGluLysGlyGlnAl 496
1216 GAGAGTTCAAGAGAAAAACAGAGACATCAGCCAGACAGAGCGAGC 1265
496 aIleGlyTyrSerGlyLeuThrGluThrGlnAlaAlaThrGlnL 513
1266 CGTGAAG.....CGGCTCCGACCGCCCTGCAGGA 1294
513 euAlaIleTyrTyrPheThrAspSer.....AlaGluLeuAspLysAsp 527
1295 GGGCCAGAGAGACCCCTGCTCGTCACAGCCAGCGCCAGCTGGAGATCGAC 1344
528 LysLeuLysAspTyrHisGlyPhe..... 535
1345 TCCCTGTTGAGGGCATCGACTTCACGTCATCACAGGGCGAGATT 1394
536 .....GLYAspMetAsnAspSerThrLeu...AlaValAlaL 547
1395 CGAGAGACTGCTCGCTCGACCTGTTCCGAGACCCCTGAGCCCTGGAGGA 1444
547 ysIleLeuValGluTyrAlaGlnAspSerAsnProGlnLeuThrAsp 563
1445 AGGCTTGCGGGAGCGCAAGCTGACAAAG.....GCCCAATTACAGGAC 1488
564 LeuAspPhePhe.....IleProAsnAsnLysTyrG1 575
1489 CTGCTGCTGCTCGGGGGCTCCACCCCGCATCCCC.....AAGTGCA 1529
575 nserLeuIleGlyThrGlnTrpHisProGlnAspLeuValAspIleLeu 592
1530 GAACCTCTCTCGAGAGCTTTCACAGCGGGCGACTGTAACAGAGACATCA 1579
592 rGmetGluAspLysLysGluValIleProValThrHisAsnLeuThrLeu 608
1580 ACCCGGACGAG.....GCTGTGGCTACGGGGCGCGCGTGTG 1614
609 ArgLysThrValThrGlyLeuAlaGlyAspArgThrLysAspPheHisph 625
1615 CAGCGGGCCATC.....CTGATGGGGGACAAAGTCCGAG..... 1647
625 eGluIleGluLeuLysAsnLysGlnGluLeuLeuSerGlnThrValL 642
1648 .....AACGTGCAGACCTGCTCTGCTGAGCTGG 1678
642 ysThrAspLysThrAsnLeuGluPheLysAspGlyLysAlaThrIleAsn 658
1679 CTCCCGCTGCTGGGCTGTGAGAGCGCGAGCGGTGATGATGACCTG 1728
659 LeuLysHisGlySerLeuThrLeuGlnGlyLeuProGluGlyTyrSe 675
1729 ATCAAGGCAATCTCACCATCCCAACAG.....CAGACGCAAGAT 1769
675 rTyrLeuValLysGluThrAspSerGluGlyTyrLysValLysValAsn 692
1770 CTTCACCACTCTCGACACCAACCCGGGGTCTGATCCAGGTAGC 1819
692 erGlnGluValAlaAsnAlaThrValSerLysThrGlyIleThrSerAsp 708
1820 AGGGCGAGAGGGCC.....ATGCACAAAGAC 1845
709 GluThrLeu.....AlaPheGluAsnAsnLysGluProValValProTh 723
1846 AACATCTGTGGGGCGCTTCGAGCTGACGCGCATCCCTCGCGCCCGCAG 1895
723 rGlyValAspGln 727
1896 GGGCGTGGCCCTAG 1908
seq_name: /cgml_8/gcdata/geneseq/NA1999.DAT:X77593
seq_documentation_block:

```

```

ID X77593 standard; DNA; 3945 BP.
AC X77593;
XX
AC 09-AUG-1999 (first entry)
XX
DE S. aureus Sdr DNA.
XX
KW Fibrinogen-binding protein; alpha chain; ClfB; SdrC; SdrD;
KW SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical;
KW treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis;
KW extracellular matrix; vascular graft; vascular stent; vaccine;
KW intravenous catheter; artificial heart valve; cardiac assist device;
KW antibacterial; ss.
XX
OS Staphylococcus aureus.
XX
PN MO9927109-A2.
XX
PD 03-JUN-1999.
XX
XX 25-NOV-1998; 98MO-US25246.
XX
PR 31-AUG-1998; 98US-0098427.
PR 26-NOV-1997; 97US-0066815.
XX
PA (EIDH/) EIDHIN D N.
PA (FOR-) FORAS T/A BIORESEARCH IRELAND.
PA (FOST/) FOSTER T J.
PA (HOOK/) HOOK M A O.
PA (INH-) INHIBITEX INC.
PA (JOSE/) JOSEFSSON E.
PA (PART/) PARTI J M.
PA (PERK/) PERKINS S E.
XX
PI Eidin DN, Foster TJ, Hook MAO, Josefsson E, Partl JM;
PI Perkins SE;
XX
DR WPI; 1999-357844/30.
XX
DR P-PDB; Y08642.
XX
PT Staphylococcus aureus fibrinogen-binding proteins for treating
PT septicemia, osteomyelitis, mastitis or endocarditis
XX
PS Claim 10; Fig 8; 143pp; English.
XX
XX
CC This invention describes novel Staphylococcus aureus fibrinogen-binding
CC proteins that bind both the alpha and beta fibrinogen chains. The
CC proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and
CC SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere
CC to medical devices, binding proteins that bind both the alpha and beta
CC fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as
CC competitive inhibitors to block this binding. Antibodies against ClfB,
CC SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding.
CC The proteins of the invention can be used in a pharmaceutical composition
CC for the treatment of Staphylococcus aureus infection e.g. septicemia,
CC osteomyelitis, mastitis or endocarditis or to inhibit the binding of
CC S. aureus to the extracellular matrix. The proteins or their fragments
CC may be used to coat a medical device to reduce the S. aureus infection of
CC an indwelling medical device, especially where the medical device is
CC selected from the group consisting of vascular grafts, vascular stents,
CC intravenous catheters, artificial heart valves, and cardiac assist
CC devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or
CC encoding gene may be used as a vaccine. The DS (aspartate serine) repeat
CC region or a gene encoding it may be used as an identifying probe for the
CC identification of genes and encoding proteins from Staphylococcus aureus
CC (other than ClfA). S. hemolyticus, S. lugdunensis, and S. schlerferi.
CC The proteins of the invention have antibacterial activity.
XX
SQ Sequence 3945 BP; 1617 A; 605 C; 714 G; 1009 T; 0 other;

```

alignment\_scores:



[illegible]

```

2958 TTCCTATGGTTTAACACACAGCGTCATTAAAGTCAGCATTAACATGA 3007
620 hLysAspPheHisPheGluIleGluLeuLysAsnAsnLysGlnGluLeu 636
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3008 CATTAAGACAGTGGTTTC.....TATTAACACCAAAATATAGTTTA 3048
637 .....LeuSerGlnThrValLysThrAspLys 646
3049 GGTGATTATGTTGGTTACGACAGTAATTAACGCGCAACAAAGTTTCAC 3098
646 rAsnLeuGluPheLysAspGlyLysAlaThrIleAsnLeuHisGlyG 663
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3099 TGAATAAGGTATCAAGATGTTAAAGTACTTATTATTAAGTAAAGGCG 3148
663 LuSerLeuThrLeuGlnGlyLeuProGlnGlyTyrSerTyrLeuValLys 679
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3149 AAGTATATGACACACTAAACAGATGAATGTAATGTAATACTGCTTGAT 3198
680 GluThrAspSerGlnGlyTyrIlyValLysValAsnSerGlnGluVal 696
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3199 AATTAGATAGCGGTAAATACAAAGTTATTTTT.....GAAAAGCCTGC 3242
696 aAsnAlaThrValSerLysThrGlyIleThrSerAspGluThrLeuAla 713
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3243 TGCGTTTACACAAACAGGTACAAATACACTGAGATGATTAAGATCAG 3292
713 heGluAsnAsnLysGluProValValProThrGlyValAspGlnLysIle 729
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3293 ATGGTGGCAAGTGCATACCAATTCAGATCATGATTCACACTT 3342
730 ...AsnGlyTyr 732
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3343 GATTAATGCGCTAC 3354

seq_name: /cognl_8/gcgdata/geneseq/geneseqn/NA1999.DAT:X20228
seq_documentation_block:
ID X20228 standard; DNA; 6168 BP.
AC
AC X20228;
XX
XX
XX
XX 20-APR-1999 (first entry)
DE
XX Enterococcus faecalis gene EF123.
XX
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis,
KM detection; attenuation; antigenic; ss.
OS Enterococcus faecalis.
XX
XX WO9850554-A2.
XX
XX 12-NOV-1998.
PD
XX
XX 04-MAY-1998; 98WO-US08959.
XX
XX 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
XX WPI; 1999-070095/06.
XX
XX P-PSDB; Y00238.
XX
XX New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines
XX for prevention or attenuation of Enterococcus infection
XX
XX Claim 1; Page 222-224; 304pp; English.
XX

```

The present sequence represents a gene isolated from *Enterococcus faecalis*. The present invention describes genes, proteins and antigenic polypeptides isolated from *E. faecalis*. The proteins can be used in vaccines for preventing or attenuating an infection caused by a member of the *Enterococcus* genus in an animal. They can also be used for detecting *Enterococcus* antibodies in a sample. The nucleotide sequences can be used for detecting *Enterococcus* nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of *E. faecalis* protein activity.

Sequence 6168 BP; 2097 A; 1176 C; 1389 G; 1506 T; 0 other:

# alignment\_scores:

Quality: 149.00 Length: 935  
Ratio: 0.369 Gaps: 44  
Percent Similarity: 43.209 Percent Identity: 19.037

## alignment\_block:

US-09-494-297-2 x X20228 ..

Align seg 1/1 to: X20228 from: 1 to: 6168

```

19 LeuSerLysAsnSerLysArgPheThrValThrLeuValGlyValPheLe 35
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 CTGAGAAAACAGCCATGATGAGACACATATTTCACGCGCCCATTT 3410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
35 UMetIlePheHlaLeuValThrSerMetVal..... 45
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3411 CCAATTGTGGACCAACGCTAAACACAACTCTACGTGAAGTACAGTAG 3460
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
46 .....GlyAlaLysThrValPheGly..... 52
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3461 ATGCCACCGGGGTTATCATCTTGCTGGTCCACAGGCAATTAATT 3510
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
53 LeuValGluSerSerThrProAsnAlaIleAsnProAspSerSerSergI 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3511 TTGGTGAGACAAAAGCACCA..... 3531
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
69 uTYrAyrTrpTyrGlyTyrGluSerTyrValArgGlyHisProTyrTyrL 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3532 .....GAAAGCTATACAGTTTCGACGAATTAAGCTA 3562
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
86 YsgInPheArgValAlaHisAspLeuArgValAsnLeuGluGlySerArg 102
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3563 AAGGCCAGCATTTACTATTGATGAGAACTTACGCCGAAGGACACAA 3612
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
103 SerTyrGlnValTyrCysPheAsnLeuLysLysAlaPheProLeuGlySe 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3613 CCA...ACCATTTATTAACAGATGTCATTAAGTATTTTAAAGAAAAT 3659
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
119 RAspSerSerValLysLysTrp.....TyrLysLysHisAspG 132
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3660 GGATGAGAAAGGTTAAAGTTAGTCAATGCTCGCTTAAATTAAGCATG 3709
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
132 LysIleSerThrLysPhe.....GluAspTyrAlaMetSerProArg 145
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3710 CCGTAACACCGCCCTTACTTCATTGGAGAGAGTTCCCTTGCCCGCAT 3759
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
146 IleThrGlyAspGluLeuAsnGlnLysLeuArgAlaValMetTyrAsnG 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3760 CGAACCC..... 3765
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
162 YHisProGlnAsnAlaAsnGlyIleMetGlu.....GlyLeuGluPro. 176
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3766 .....AACCGAATGCGCCATTAGAGTGATAGTTTAAACACAG 3805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
177 ..LeuAsnAlaIleArgValThrGlnGluAlaValTrpTyrTyrSerAsp 192
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3806 GCGCTTATTCAGTTTCAGAAATCGAAGCACCAGACGCTATCTTTAGAC 3855
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
193 AsnAlaPro..... 195

```

```

3856 ACAGACCCCAAGATTCATCTGACACAAATAACGACGCAAAATTCG 3905
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
196 .....IleSerAsnProAspGluSerPheL 204
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3906 TGATGTTTCATGTCAAAATGCTTATTAACCAAGGTTCTGCTGAACATAATTA 3955
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
204 YsArgGluSerGluSerAsnLeuValSerThrSerGlnLeuSerLeuMet 220
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
3956 AAAAGACCAAGCAGCATTCATTCATGACAGTGTGAAATTTTCAGTCCTT 4005
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
221 .....ArgGlnAlaLeuLysGln.....LeuIleAspProAsnLeu 232
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4006 GACACACACAGCAGACAGCATTCGAGAACACTTAGTTGCGATCAACAGG 4055
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
232 uAlaThrLysMetProLysGlnValProAspAspPheGlnLeuSerIleP 249
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4056 AAAAGTCACAGTACGAGGATTTTACCCCAAGAAATATCAATTT..... 4098
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
249 heGluSerGluAspLysGlyAspLysTyrAsnLysGlyTyrGlnAsnLeu 265
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4099 .....GTGGAACCAAGCGCCA..... 4116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
266 LeuSerGlyLeuValProThrLysProProThrProGlyAspProPr 282
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4117 ...GCAAGGTACCTTTTAAACACTGAACCAAGCTTTTCACGATTCGACG 4163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
282 oMetProProAsnGlnProGlnThr..... 290
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4164 AAGCGATCGGGGCAACACACACAGCTTATACAGCGCTAATTGTTTA 4213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
291 .....ThrSerValLeuIleArgLysTyrAlaIleGlyAspTyr 303
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4214 ACTATCAAGGACAGCGCTAATTAATCAAAAAGATGAGTGA..... 4257
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
304 SerLysLeuLeuGluGlyAlaThrLeuGlnLeuThrGlyAspAsnValAs 320
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4258 ...CACTATTAAAGTGTGCGACATTTAAAGTCTTGATCGGAAGGAGA 4304
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
320 nSerPheGlnAlaArgValPheSerSerAsnAspIleGlyAlaArgL 336
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4305 AACCATTCACAAAGCGCTTG...ACGACAAATATCAAGGGGAATTTGTT 4351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
337 .....GluLeuSerAspGlyThrTyrThrLeuThrGluLeuAsnSerPro 351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4352 CAGAGCACTTAGCCCGCAGAAATATCGCTTGTAGAACCAAGCGCCA 4401
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
352 AlaGlyTyrSerIle...AlaGluProIleThrPheLysVal..... 364
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4402 ACAAGGCTATTATTAATTAACACGCGCAGTCCCATTTGAAATTCCTGAGAA 4451
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
365 ..GluAlaGlyLys..... 368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4452 AATGCTGTGTAACACAGCGCTCGTGTGCTAGTACAACTTTGTGAGTT 4501
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
369 .....ValTyrThrIleIleAspGlyLysGlnIleGluAsnProAsn 382
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4502 ACAAGGCGGCTTCCAAATCGTGAAGAAAGATAGCGCAGACCAACCATTA 4551
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
383 LysGluIleValGluProTyrSerValGluAlaTyrAsnAspPheGluG 399
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4552 GCAAGTCTGCTT.....TTTGAATTAAT...GATCACAAATAA 4586
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
399 uPheSerValLeuThrThrGlnAsnTyrAlaLys..... 410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4587 ACAATCATTTAGGATTACACACAGCAAGCTGCAAGAGATGCAAAATTTATCT 4636
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
411 .....PheTyrTyrAlaLysAsnLys..... 417
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4637 TTAGAGACTTGGCCCGCAGTACATATATTAACAAGAAATCAAGACCA 4686
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
418 .....AsnGlySerSerGlnValValTyrCys..... 426
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

4687 AATTACAGATGGCGAGATTATATTATTCCTGAATTAGTAAAGT 4736
427 ... PheAsnAlaSplLeuLysSerProProAspSerGluAspGlyL 442
      :::::::::::::::::::: |||:::
4737 AGAAATTCGTGTGATTTCAAAGGTGATCCGAGATT 4773
442 ysthmetThrProAspPheThrGlyGluValLysTyrThrHisIle 458
      ||| |||
4774 .....TTCCAATTAGG.....GCCCTCGCAAAATTTC 4800
459 AlaGlyArgAspLeuPheLysTyrThrValLysProArgAspThrAspPr 475
      ||||| ::::::::::::::
4801 AAAGAGCGCGCGCTTTAGAAAATTGATGCCAATGCCAACCACCTCC 4850
475 AspThrPheLeuLysHisIleLysLysValIleGluLysGlyTyrArg 491
      | :::::::::::::: |||:::
4851 AGCAACGATTTTAAATTGTATCGA.....ATCGAAACCGGGGAAAAA 4894
492 .....GluLysGlyIleAlaIle..... 497
4895 TCTTTGAAAGAGAACTACTGCTGAAAAAGATGGTTCATTGCTATGGAG 4944
498 .....GluTyrSerGlyLeuThrGluThrGlnLe 507
4945 GATTAGGTGCTGTAGCTATGATTAATGATGACTGATGCACAGGATGG 4994
507 uArgAlaIleThrGlnLeuAlaIleTyrTyrPhe.....T 519
      :::::::::::::: |||:::
4995 CTATATGCTCAATTAACACACCATTTATTTTGTAGTGAAGAGAATTCAA 5044
519 hrAspSerAlaGluLeuAspLysAspLysLeuLysAspTyrHisGly... 534
      :::::::::::::: |||:::
5045 ATGATTAACAACAACACTAGATGAGTTAGAGTTTGAATATATCAAGCGAA 5094
535 ...PheGly.....AspMetAsnAspSerThrLeuAlaValAlaLys 547
      ::||| :::: |||::|
5095 GTATGGAGCTTAAGTCACAGCAGCAAGCTTAGCGGCGCAGT 5144
547 sIleLeuValGluTyrAlaGlnAspSerAsnProGlnLeuThrAspL 564
      :::: |||::: |||::: |||:::
5145 TTTTGCAATTTTACATGCGCATGAGCAGAAATCACGCCCAAGGTCACCGA 5194
564 euAspPhePheIleProAsnAsnAsnLysTyrGlnSer..... 576
      ::|||::: :::: |||:::
5195 TAACATTCCTGTAATCGTCAGAGAGAAAAAGTTCTGAATTAACAACGAT 5244
576 ..... 576
5245 AAGACTGGCGAATTTTACGCTAAAGCGCTAAATGAGGCAATTACGTTT 5294
577 .....LeuIleGlyThrGlnThrHisP 584
      ::|||::: |||::: |||:::
5295 AGTGAAACGAAACACCAACACAGCGCTATCTGTGACACACAGCTACATC 5344
584 roGluAspLeuValAspIleIleArgMetGluAspLysLysGluValIle 600
      |||::: |||::: |||::: |||:::
5345 CA.....TTTGATGTAAACCGCCCAATTAGAGAAAAAGCAGCCCAATT 5385
601 ProValThrHisAsnLeuThrLeuArgLysThrValThrGlyLeuAlaG 617
      :::: |||::: |||:::
5386 GCTTAGCGCATCTTATCAATTATCAAGAACTGCT..... 5421
617 yAspArgThrLysAspPheHisPheGluIleGluLeuLysAsnAsnLysG 634
      :::: |||::: |||:::
5422 .....CAATTAAACCAAGAAAAAGAAACAG 5446
634 lngIleLeuLeuSerGlnThrVal.....LysThrAspLysThrAsnLeu 648
      ||| |||::: ::||| |||::: |||:::
5447 GTGAAGCATTTGGCAGGTGGGTGTTTAAGTCATTGATGAAGAAACAGG... 5493
649 GluPheLysAspGlyLysAlaThrIle...AsnLeuLysHisGlyLys 664
      ::|||::: |||::: |||::: |||:::
5494 CAATCGTAGATGAGCAACCAATCTGATGTGTGACAAAGCAAGCAAAAGT 5543
664 rLeuThrLeuGlnGlyLeuProGluGlyTyr..... 674
      :::: ||| |||
5544 CATTCGAAAAACCTTACACCGGAGACGTATCTTTGTGGAGACACAAAG 5593
675 .....SerTyrLeuValLysGluThrAspSerGluGlyTyrLys... 687
      ||| |||::: |||::: |||:::
5594 CGCCACCTAGCTATCTTCTTAATGAAACGCCAAGCGCAAGCTTACGATT 5643
687 ..... 687
5644 GCCAAAGACAAACCAAGCAACACGACCTGTGTAAGTAAAGCACCCTT 5693
688 .....ValLysValAsnSerGlnGluVal 695
5694 TATTAAATTACCAAGTGTGTGCCAAGCTGTGATAAATTGATCAGCAAAAGA 5743
695 AlaAlaAsnAla..... 698
5744 ATGCCTTAGCAGGTGCTGAATTTAAAGTGACAGATGCCAGACAGGCGAA 5793
699 ThrValSerLysThrGlyIleThrSerAspGluThrLeuAlaPheGluAs 715
      ||| |||::: |||::: |||:::
5794 ACTGTCGCTGCTCATTTACGTTCTGACAAACCAAGGTTAGTTCAAGTGA 5843
715 nasnLysGluProValValProThrGlyValAspGlnLysIle...AsnG 731
      :::: ||| |||::: |||:::
5844 CCACCTTACMACCGAAGAAATATATACCTTTGTGTAACAAACACCGCGATG 5893
731 LysTyr 732
5894 GTTAC 5898

```

